Pseudouridine synthase 7 is an opportunistic enzyme that binds and modifies substrates with diverse sequences and structures

Meredith K. Purchal[1], Daniel E. Eyler[2], Mehmet Tardu[2,4], Monika K. Franco[1], Megan M. Korn[2], Taslima Khan[1], Ryan McNassor[2], Rachel Giles[2], Katherine Lev[1], Hari Sharma[2], Jeremy Monroe[2], Leena Mallik[2,3], Markos Koutmos[1,2,3]* and Kristin S. Koutmou[1,2]*

[1]Program in Chemical Biology, University of Michigan, Ann Arbor, Michigan
[2]Department of Chemistry, University of Michigan, Ann Arbor, Michigan
[3]Department of Biophysics, University of Michigan, Ann Arbor, Michigan
[4]currently affiliation, Department of Biology, Istanbul University, Istanbul, Turkey

*Corresponding authors: Kristin Koutmou and Markos Koutmos
Email: kkoutmou@umich.edu, mkoutmos@umich.edu

Running Title: Pus7 structure and substrate selection mechanism (include on first page)

Keywords: pseudouridine, RNA modification, Pus7, TruD, structure (5-10 keywords, find lists online)

ABSTRACT

Pseudouridine (Ψ) is a ubiquitous RNA modification incorporated by pseudouridine synthase (Pus) enzymes into hundreds of non-coding and protein coding RNA substrates. Here, we determined the contributions of substrate structure and protein sequence to binding and catalysis by pseudouridine synthase 7 (Pus7), one of the principal mRNA modifying enzymes. Pus7 is distinct among the Pus proteins because it modifies a wider variety of substrates and shares limited homology with other pseudouridine synthase family members. We solved the crystal structure of *Saccharomyces cerevisiae* Pus7, detailing the architecture of the eukaryotic specific insertions thought to be responsible for the expanded substrate scope of Pus7. Additionally, we identified an insertion domain in the protein that fine-tunes Pus7 activity both *in vitro* and in cells. Our data demonstrate that Pus7 preferentially binds substrates possessing the previously identified UGUAR (R = purine) consensus sequence, and that RNA secondary structure is not a strong requirement for Pus7 binding. In contrast, the rate constants and extent of Ψ incorporation are more influenced by RNA structure, with Pus7 modifying UGUAR sequences in less structured contexts more efficiently both *in vitro* and in cells. Although less structured substrates were preferred, Pus7 fully modified every tRNA, mRNA and non-natural RNA containing the consensus recognition sequence we tested. Our data suggest that Pus7 is a promiscuous enzyme, and lead us to propose that factors beyond inherent enzyme properties (e.g. enzyme localization, RNA structure, and competition with other RNA-binding proteins) largely dictate Pus7 substrate selection.

SIGNIFICANCE STATEMENT

Pseudouridine is among the most abundant RNA modifications. We present a framework for conceptualizing how eukaryotic pseudouridine synthases select their substrates. This work reveals the structure of yeast pseudouridine synthase 7 (Pus7) and presents cell-based and biochemical investigations of enzyme binding and activity. We demonstrate that Pus7 interacts promiscuously with RNAs containing UGUAR sequences. Our observations raise the question of why these enzymes only modify < 5% of UGUAR sequences in the transcriptome, suggesting that factors beyond inherent enzyme properties - such as protein localization, local RNA structure and mRNA-protein interactions - principally shape Pus7 substrate selection. These findings support a role for Pus7 in providing cells with a mechanism to rapidly alter protein synthesis in response to cellular conditions.

INTRODUCTION

Post-transcriptional modifications to the four standard RNA nucleosides increase the structural and functional complexity of RNAs. The C5-glycosidic isomer of uridine, pseudouridine (Ψ), is incorporated into multiple RNA species including transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and eukaryotic messenger RNAs (mRNAs) (Fig. 1A). While Ψ has been studied in the context of non-coding RNAs for decades, the significance of Ψ in mRNAs is less well understood. Like all mRNA modifications Ψ has the potential to affect every step in the life-cycle of an mRNA (1). In line with this idea, mRNA pseudouridylation has been implicated as a regulator of human alternative splicing, yeast protein synthesis and toxoplasma mRNA metabolism (2-4). Nonetheless, despite its prevalence, how cells select mRNAs for Ψ modification, and the impacts of individual mRNA Ψ-sites on biological processes remain to be established.

Ψ is installed into RNAs in all organisms by pseudouridine synthases (Pus). This large class of enzymes is categorized into six families: TruA, TruB, RluA, RsuA, TruD and Pus10 (SI Appendix Fig. S1). Despite the observation that many Pus enzymes are not required for cellular viability in non-stressed conditions, wild-type cells outcompete Pus-deficient cells suggesting that these enzymes confer a fitness advantage (5-7). Consistent with this, Pus proteins markedly enhance cellular fitness under heat shock (8-11). Furthermore, mutations in the human TruA, TruB and TruD family members (Pus1, Pus4 and Pus7, respectively) are linked to inherited diseases such as mitochondrial myopathy and sideroblastic anemia (MLSA) and intellectual disabilities (12-15). In addition to their enzymatic roles, there is emerging evidence that some Pus proteins can have alternative functions in the cell - acting as a tRNA folding chaperones or prions (16-18). One of the next horizons for the pseudouridine synthase field will be deconvoluting the contributions of each of the varied Pus activities (non-coding RNA- and mRNA-modifying, or non-enzymatic) to gene expression.

Bacterial Pus family members have been studied for decades and their structures, chemical mechanisms and modes of RNA target selection are well characterized (19, 20). However, it is unclear why some eukaryotic Pus enzymes exhibit an increased substrate scope, modifying hundreds of mRNAs in addition to their traditional non-coding targets (21, 22). The enzymes responsible for catalyzing the bulk of Ψ incorporation into eukaryotic mRNAs are Pus1, Pus4 and Pus7 (23, 24). Recent transcriptome wide Ψ-mapping and reporter studies indicate that mRNA secondary structure and sequence contribute to substrate selection by human Pus1 and Pus4 (25, 26). Pus7 is distinct from Pus1 and 4, as it shares ~10% sequence identity with these enzymes and incorporates Ψ into a particularly diverse set of RNAs (tRNA, tRNA fragments, snRNA, rRNA, and mRNA) (23, 24, 27-29). Pus7 has been shown to preferentially incorporate Ψ into UGUAR (U = U converted to Ψ, R = A or G) consensus sequences (23, 24, 28). However, this sequence motif does not explain how Pus7 selects its substrates because only a limited set (< 5 %) of the UGUAR sequences present in RNAs are Pus7 targets.

To establish the determinants of Pus7 substrate selection, we first solved a crystal structure of *S. cerevisiae* Pus7. Comparing the yeast and human Pus7 structures with their prokaryotic TruD counterparts, we noticed a large eukaryotic-specific insertion (insertion domain I, ID-I) sitting atop the active site in two distinct conformations. Both of the observed conformations for this domain

are well-positioned to potentially interact with incoming RNA substrates. Our extensive biochemical and cell-based studies of Pus7 enzymes lacking ID-I ($Pus7_{\Delta IDI}$) support this supposition. We also evaluated the impact of RNA sequence and structure on Pus7 substrate binding and modification. Our results indicate that Pus7 preferentially associates with RNAs containing a UGUAR consensus sequence, but only minimally discriminates between binding different RNA structures. The enzyme has greater degree of selectivity during catalysis than binding, with the rate constant and extent of pseudouridylation increasing on RNAs with less predicted structure both *in vitro* and in cells. Nonetheless, despite these preferences, Pus7 modifies all of the UGUAR containing substrates we presented it *in vitro*. Our observations lead us to propose an opportunistic model for Pus7 substrate selection. In this model, Pus7 rapidly samples RNAs, binding and modifying any UGUAR sequence it can access for long enough. Our model is supported by observations that UGUAR sequences in structured contexts are modified more rapidly and efficiently at elevated temperatures, when RNA structures are destabilized. This extends the current framework for thinking about Pus enzymes by proposing that protein compartmentalization, local RNA structure, and mRNA-protein interactions are among the most substantive determinants for Pus7 substrate discrimination.

RESULTS

S. cerevisiae Pus7 structure reveals flexibility in the architecture of the eukaryotic insertion domains

Pus7, like other TruD family members, shares little sequence identity with the four other pseudouridine synthase families (TruA, TruB, RluA, RsuA)(30). Available structures of TruD from *E. coli* [PDB:1SI7, 1ZSW, 1SB7], *Methanosarcina mazei* [PDB:1Z2Z], and Pus7 from humans [PDB: 5KKP] show that the core of the enzyme has a V-shape formed by two conserved domains - the catalytic PUS domain and the TRUD domain (31-33). The structure of the PUS domain is highly conserved across all pseudouridine synthases (34), with the invariant catalytic aspartate falling on a β-strand that lies across the center of this domain. The TRUD domain is a mixed αβ fold conserved across all TruD homologs (30-33). This domain is oriented adjacent to the PUS domain, and together they form a catalytic cleft (V) lined with positively charged residues. The TRUD and PUS domains are connected by an extensive loop-rich region at the base of the cleft where the active site sits. Structural comparison of the five biological assemblies present in the three *E. coli* TruD structures demonstrates the ability of these core domains to flex towards each other by 18° (32). This motion could be facilitated by the hinge-like loop-rich region connecting the domains, and potentially help the enzyme to "clamp" down on RNA substrates.

Archaeal and eukaryotic TruD homologs contain large insertions at specific sites in the TRUD domain (33). Eukaryotic homologs possess two additional unique insertions decorating the catalytic domain (31-33). The largest of these eukaryotic-specific insertions, insertion domain I (ID-I), is a small domain first visualized in the structure of human Pus7 (PDB: 5KKP). This domain is conserved in yeast Pus7 (SI Appendix Fig. S2) (30). Eukaryotic members of the TruD family modify a much larger range of RNAs than their bacterial counterparts, which only modify U13 on $tRNA^{Glu}$ (30). Given their conservation and the expanded substrate selection observed, it stands to reason that structural features (insertions) unique to the eukaryotic enzymes might play a crucial role in the recognition and modification of their additional substrates.

To that end, we sought to develop a structural basis for understanding how eukaryotic insertions can contribute to Pus7 substrate selection and modification. We used a combination of molecular replacement (using human Pus7 as a model) and single-wavelength anomalous dispersion phasing to obtain a crystal structure of *S. cerevisiae* Pus7 to 3.2 Å resolution (Fig. 1B, SI Appendix Table S1 and Fig. S2) (PDB:7MZV). The core of the Pus7 structure strongly resembles that of TruD (RMSD =3.74 Å based on 144 atoms), adopting the characteristic V-shape conformation of the PUS and TRUD domains (Fig. 1C, SI Appendix Fig. S2). The active site sits near the bottom of the cleft adjacent to the hinge region where the two domains interface. The universally conserved catalytic aspartate (D256) resides on a loop between β3 and β4 and is surrounded by residues conserved within the TruD family (K61, F67, E71, N305 and F307) (SI Appendix Fig. S3), and is ~ 4 Å closer to the conserved active site residues in Pus7 than in TruD (SI Appendix Fig. S2). Such positioning is consistent with the location of the catalytic Asp in the co-crystal structure of TruB with an RNA (PDB:1K8W) (32). The most notable structural difference between Pus7 and TruD is the presence of the three eukaryotic specific insertions that decorate the enzyme's V-shaped core [insertion domain I (ID-I, aa 75 – 215), II (Ins-II, aa 365 –

443), and III (Ins-III, aa 594 – 634)] (Fig. 1B). These insertions are connected to the catalytic domain by flexible linkers.

A structural alignment of yeast and human Pus7 reveals that the three insertions adopt similar folds and are present at equivalent locations in both structures (SI Appendix Fig. S2). ID-I folds into an extensive domain that is anchored to the top-exterior (side opposite the active site) of the PUS domain by flexible linkers. ID-I is observed in two distinct orientations above the active site (SI Appendix Fig. S2). Ins-III, the smallest insertion, is located on the exterior of the catalytic domain near the hinge region. Ins-II is a helical bundle perched atop the TRUD domain, and together, ID-I and Ins-II elongate the cleft between the PUS and TRUD domains. These insertions are ideally positioned to potentially interact with RNA substrates. Comparison of human and yeast Pus7 structures demonstrates that ID-I and Ins-II can adopt different orientations relative to the core of the protein. In yeast Pus7, ID-I extends laterally away from the protein, while in human Pus7 it is slightly rotated to be positioned directly above the active site. This suggests that ID-I can move as a rigid body, both swiveling laterally (akin to a flag on a pole) over the top of, or away from the active site. Ins-II is also found in slightly different orientations in each structure, hinting to its rotational and translational freedom of motion. Ins-II in yeast Pus7 is pivoted away from the active site, but angled in towards the catalytic cleft in human Pus7. Though the physiological relevance of these orientations is not immediately apparent, the freedom of motion and inherent flexibility could enable Ins-II or ID-I to play a part in recognizing regions of the substrate distal to the site of modification.

To determine if ID-I shares homology with known RNA binding motifs, we performed a protein structure comparison search with the DALI server (35). Using ID-I from both the yeast and human Pus7 structures, our results indicate that ID-I shares strong structural homology with the RNA binding R3H domains. The strongest match was to a Poly(A)-specific ribonuclease domain (PDB: 2a1s, Z = 6.3, %ID =21) that utilizes an R3H motif to bind single stranded nucleic acids. This motif contains an invariant arginine separated from a conserved histidine by three residues. Sequence alignments demonstrate that these residues are present and conserved in ID-I (SI Appendix Fig. S3). Additionally, the structure of the R3H motif is characterized by two α–helices packed against a three-strand β-sheet, which aligns well with the structures of ID-I in yeast and human Pus7 (SI Appendix Fig. S4). ID-I's predicted ability to interact with RNAs and its stark positional variance relative to the active site suggest that ID-I may serve as an attenuator of RNA binding.

Pus 7 enhances S. *cerevisiae* viability under translational stress

Pus7 enhances cellular fitness under temperature stress and modifies a wider variety of mRNA targets under heat shock (24). Given the recently discovered links between Pus7 mutations and neurological defects, we hypothesized that Pus7 is important for cellular health under additional stress conditions (12). To test this supposition, we compared the growth of wild-type and *pus7*Δ *S. cerevisiae* cells under 15 different conditions: multiple temperatures (22^{o}C, 30^{o}C, 37^{o}C), elevated salt concentrations (NaCl, $MgCl_2$), varied pH (pH 4.5 and 8.5), carbon sources (glucose, sucrose, galactose), proteosome stress (MG132) and translational stress (puromycin, paromomycin, cycloheximide, hygromycin). Cell growth was assessed by spot-plating on solid

media and growth curves in liquid media (SI Appendix, Figs. S5, S6). *pus7*Δ cells do not exhibit a growth defect relative to wild-type cells in YPD media at 22ºC or 30ºC, but are sensitive to increased temperature (37ºC) as previously reported (11). We did not observe any carbon source-dependent growth changes between the wildtype and the knockout cells. The *pus7*Δ strain has a slight sensitivity to high concentrations of NaCl, but not $MgCl_2$. This is consistent with high-throughput studies that identified *pus7*Δ as one of ~300 yeast knockouts that are more sensitive than wild-type cells to hyperosmotic (1M NaCl) stress (36).

Translation inhibitors had the largest impact on *pus7*Δ growth relative to wild-type cells. *pus7*Δ cells were more sensitive to puromycin, cycloheximide and hygromycin, and exhibit a decreased sensitivity to paromomycin (SI Appendix Figs. S5, S6). This makes sense as Pus7 modifies both tRNAs and mRNAs, which could impact translation. To test hypothesis, we analyzed available ribosome profiling data sets for *pus7*Δ (37, 38). We observe that ribosome occupancy is increased on Pus7-targeted mRNA codons in *pus7*Δ cells (SI Appendix Fig. S7), consistent with recent reports that mRNA pseudouridinyltion slows translation elongation (3, 39). Together, our results indicate that Pus7 is likely to be particularly important when cellular translation is under stress.

Conserved Pus7 active site residues enhance RNA modification

Our crystal structure revealed that the Pus7 active site is similar to that of TruD, and suggests that ID-I might be positioned to contribute to enzyme function. We next wanted to test if conserved Pus7 active site residues and ID-I enhance the ability of the enzyme to modify a reported mRNA substrate, CDC8 (24). To accomplish this, we measured the single turnover rate constants (k_{obs}) for Ψ incorporation into a 61nt long region of CDC8 by wild-type and mutant enzymes (D256A, K61A, E71A, F67A, H161A, N305A, F307A, F307Y and $Pus7_{\Delta ID\text{-}I}$) (Fig. 2, Table 1). In these experiments the Pus7 enzymes (2-10 μM) were incubated with ^{3}H-labeled CDC8 RNA (< 100 nM) and ^{3}H release upon conversion of U to Ψ was monitored at discrete time points (40). As expected, mutation of the catalytic D256 residue to alanine abolishes Pus7 activity, with no Ψ formation observed after 16-hours (41). Alanine substitution of the nearby active site residues K61, F67 and E71 reduced the k_{obs} for CDC8 modification by 40 to 200-fold, consistent with the proposed role of these residues in substrate positioning during catalysis in TruD. Mutations to the conserved active site NxF motif had larger impacts on Pus7 activity, with N305A and F307A mutants decreasing k_{obs} by up to 50,000-fold relative to the wild-type enzyme. The F307A defect was partially rescued by a F307Y mutation (400-fold reduction in k_{obs}) suggesting that the F307 base stacks with the target uridine to enhance CDC8 modification. In contrast to the active site mutants, k_{obs} is unchanged by the ID-I point mutation H161A. However, removal of ID-I truncation ($Pus7_{\Delta ID\text{-}I}$) reduced k_{obs} by 2-fold (SI Appendix Table S2). These data reveal that conserved active site residues are important for Ψ modification by Pus7, while ID-I does not have a large influence on the rate determining step for Pus7 under saturating enzyme concentrations.

Multiple Pus7 enzymes bind to unmodified and modified CDC8

To assess the contributions of active site residues and ID-I to RNA substrate binding we performed electrophoretic mobility shift assays (EMSAs) with a 5'-fluorescein labeled CDC8 and

a series of catalytically inactive Pus7 mutants (D256A mutant background). We observed a single band shift at low enzyme concentrations (< 50 nM), indicating the formation of a Pus7-CDC8 (ES) complex (Fig. 3). For all of the mutants evaluated, the singly-bound species is super-shifted when enzyme concentrations are increased (Fig. 3, SI Appendix Fig. S8). These additional bands indicate the formation of complexes that contain multiple Pus7 proteins associated with CDC8 (ESE_n complexes). We speculate that these lower-affinity binding events reflect non-specific interactions between Pus7 and the RNA phosphodiester backbone. This may be a general property of Pus enzymes, as we also observe a similar binding behavior for the pseudouridine synthase family member Pus1 interacting with Fl-CDC8 (SI Appendix Fig. S9). For both enzymes, the transition from the 1:1 complex to the n+1:1 complex occurs over a very narrow concentration range, reflecting the large number of non-specific binding sites available on each RNA. Based on these observations, we considered several binding models (SI Appendix Fig. S10), and ultimately fit our data to a simplified mechanism in which Pus7 independently binds a unique site tightly ($K_{D,app1}$ - low nM) and multiple additional sites with reduced affinities ($K_{D,app2}$ - high nM to μM) (SI Appendix, EMSA methods). Such a model is supported by stopped-flow studies with 5'-fluorescein labeled CDC8 and unlabeled D256A Pus7. At low enzyme concentrations we observed a single exponential phase, whose rate constant (k_{obs1}) is linearly dependent on enzyme concentration (SI Appendix Fig. S11). As we increase enzyme concentration, a second phase emerged (k_{obs2}) consistent with our EMSAs indicating that multiple proteins bind to Pus7 RNA targets.

We applied our binding model to obtain $K_{D,app1}$ values for D256A, D256A/K61A, D256A/F67A, D256A/E71A, D256A/H161A, D256A/N305A, D256A/F307A, and D256A/$Pus7_{\Delta ID-I}$ binding to CDC8 (Table 1). D256A binds CDC8 tightly with $K_{D,app1}$ = 60 ± 15 nM, consistent with the K_D estimated for D256A from our stopped flow assays ($k_{on,app}$ = 4.3 x 10^8 $M^{-1}s^{-1}$, $k_{off,app}$ = 35 s^{-1}, $k_{off,app}/k_{on,app} = K_D$ ~ 85 nM) (SI Appendix Fig. S10). Additional active site point mutations increased the $K_{D,app1}$ for Pus7 binding CDC8 by 2- to 8-fold relative to D256A (Table 1; SI Appendix Fig. S8). Similarly, removal of ID-I (D256A/$Pus7_{\Delta ID-I}$) increased the $K_{Dapp,1}$ for CDC8 by 2-fold. Many enzymes bind their products with different affinities than their substrates. To test if Pus7 discriminates between unmodified and Ψ-modified transcripts we measured the $K_{Dapp,1}$ values for Pus7 and $Pus7_{\Delta ID-I}$ binding to modified CDC8. We found that wild-type Pus7 bound Ψ-modified CDC8 sequences with an affinity similar to its catalytically inactive counterpart (D256A) for an unmodified CDC8 substrate (Table 1). In contrast, $Pus7_{\Delta ID-I}$ had an 8-fold weaker affinity for a Ψ-modified CDC8 than D256A/$Pus7_{\Delta ID-I}$. This indicates that Pus7 does not distinguish between substrate or product during binding, and that ID-I promotes this lack of discrimination.

ID-I influences the extent of Ψ-incorporation in full-length mRNAs *in vitro* and in cells

Our binding studies suggest that ID-I has the potential to impact substrate selection. We tested this hypothesis by evaluating Ψ-incorporation under conditions where Pus7 must distinguish between all cellular RNAs. Pus7 and $Pus7_{\Delta ID-I}$ proteins (50 μM) were incubated for 10 minutes at 30°C with 150 μg of total RNA purified from *pus7*Δ cells. The extent of Ψ-incorporation on previously reported Pus7 modified substrates (ARG5,6, BET2, TEF5, $RTC3_{(U77)}$, $RTC3_{(U288)}$,

TEF2$_{(U555)}$, TEF2 $_{(U1104)}$) was measured using CMC-RT and Ligation Assisted PCR analysis of Ψ modification (CLAP) (42). The RNAs treated with Pus7 and Pus7$_{\Delta ID\text{-}I}$ exhibited a similar range of stoichiometries (6-40% vs. 7-60%) (SI Appendix Fig. S12). While the range of Ψ-incorporation levels was similar, the presence of ID-I impacted the extent of Ψ addition on 4 of the 7 substrates we examined - with Pus7$_{\Delta ID\text{-}I}$ incorporating Ψ at lower levels than Pus7 in ARG5,6, BET2, and TEF2$_{(U555)}$, and adding more Ψ in RTC3$_{(U288)}$ (SI Appendix Fig. S13). We verified that ID-I influences how much Ψ Pus7 incorporates in cells by using CLAP to measure Ψ-levels on BET2, RTC3$_{(U77)}$, RTC3$_{(U288)}$, TEF2$_{(U555)}$ and TEF2 $_{(U1104)}$ mRNAs purified from cells containing or lacking ID-I (SI Appendix Fig. S13). Both *in vitro* and in cells, ID-I does not have a uniform effect on the substrates that we examined. Together, these data suggest that ID-I may act as a rheostat to fine-tune how Pus7 interacts with individual sequences. The idea that ID-I makes subtle contributions to Pus7 function is supported by our observation that cells expressing Pus7$_{\Delta ID\text{-}I}$ do not have a growth defect (SI Appendix Figs. S14 and S15).

Pus7 tightly binds RNAs with an array of sequences and secondary structures

The sequences targeted by Pus7 in cells often occur multiple times within a single mRNA, yet only a small subset of potential target Us are converted to Ψ. Pus1 and Pus4 are reported to use RNA secondary structure to recognize their substrates, and we wondered if RNA structure similarly dictates which uridines Pus7 modifies (25, 26). To begin asking this question, we predicted the secondary structure context of *S. cerevisae* mRNA sequences modified by Pus7 (23). For each sequence, we modeled a 100-nucleotide region surrounding the site of Ψ-incorporation. Secondary structure models were obtained using two different folding algorithms in the RNAstructure software package (free energy minimization (MFE), maximum expected accuracy (MEA)) (23, 43). These predictions indicate that in cells Pus7 modifies Us in a wide variety of structural contexts including unstructured regions, loops, bulges and helices (SI Appendix Fig. S16). We noted that multiple Pus7 consensus sequences are often present within a single targeted mRNA. Therefore, we also examined the structural context of non-modified sequences within Pus7 targeted mRNAs. Comparison of our structural models suggest that targeted uridines more commonly exist in less structured contexts than non-targeted uridines present on the same mRNAs (SI Appendix Fig. S16).

To experimentally evaluate the ability of Pus7 to interact with RNAs in a variety of structural contexts, we measured the binding of a catalytically inactive D256A Pus7 mutant to a series of 5' fluorescein labeled RNAs (Fig. 4A). These RNAs differ in both their sequence and secondary structures and include a natural tRNA target (tRNA^{Asp}), three truncated CDC8 mRNAs (CDC8-A, CDC8-B, CDC8-C) and two non-natural substrates that place the target U in different structural contexts (ST1 and ST2). D256A bound to nearly all of the RNA substrates that we tested with similar affinities (16-130 nM) (Table S3). Only the short (19 - 25 nt) CDC8-B and ST2 substrates significantly increased the $K_{D,app1}$ for D256A, though the enzyme still bound the CDC8-B with a sub-μM dissociation constant ($K_{D,app1}$ = 800 ± 320 nM) (Fig. 4, SI Appendix Table S3). These findings indicate that Pus7 has a substantial affinity for RNAs in general, and together with our stopped-flow binding data, lead us to propose that the enzyme rapidly searches for consensus sequences amidst many nonspecific binding sites (SI Appendix Fig. S17).

Pus7 can rapidly incorporate Ψ on a diverse set of RNAs

Our results indicate that PUS7 substrate binding is largely independent of the RNA secondary structures and sequences that we investigated. We wondered if RNA structural properties play a bigger role in determining the ability of Pus7 to modify substrates. To test this, we compared the single-turnover rate constants for Ψ incorporation into $tRNA^{Asp}$, CDC-8, CDC8-A/B/C, ST1 and ST2 (Figs. 4 and 5). We find that wild-type Pus7 modified the CDC8 mRNA substrates ~10-fold faster than $tRNA^{Asp}$ ($k_{obs,tRNA}$ = 0.009 s^{-1} ± 0.0005, $k_{obs,cdc8}$ = 0.99 s^{-1} ± 0.1). The $k_{obs,tRNA}$ value that we measured is slower than previous studies of *E. coli* TruA, TruB and RluA Pus enzymes which have rate constants between 0.1 - 0.7 s^{-1} on their non-coding targets (44-46). As expected, no Ψ is incorporated into substrates when the target uridine is mutated to a cytidine ($tRNA^{Asp}_{NT}$, $CDC8_{NT}$). While the truncation of CDC8 (CDC-A/B/C) does not alter the rate constant for pseudouridylation, we find that Pus7 incorporates Ψ 200- to 4000-fold more slowly into the shorter non-natural UGUAG containing RNA sequences (ST1, ST2). We anticipate that the reduced k_{obs} for ST2 reflects its weak binding of Pus7 (Fig. 4, SI Appendix S8). Notably, the removal of ID-I ($Pus7_{\Delta ID-I}$) partially recovered enzyme activity on ST1 (increasing $k_{obs,ST1}$ by 2-fold) suggesting that this domain may help to serve as a gatekeeper for substrate selection (Fig. 5). Together, our data demonstrate that Pus7 is capable of fully modifying any substrate containing a UGUAG target sequence, regardless of context. However, it does convert uridines to Ψ more quickly when they are present in regions of RNAs > 25 nt in length predicted to be flexible single stranded.

Pus7 activity towards Us predicted to be in structured regions is enhanced at increased temperatures

We further explored our observation that Pus7 appears to be more active on less structured UGUAR sequences by measuring the rate constants for Ψ-incorporation into the ST1 substrate at elevated temperatures (37°C, 42°C, 50°C) where the structural stability of the RNA is reduced and the molecule should be more dynamic. If, as we hypothesize, base-pairing limits Pus7 activity towards substrates, then we anticipate that heating samples should increase pseudouridylation. Indeed, we find that $k_{obs,ST1}$ is increased by 18-fold between 30°C and 42°C (Fig. 6A, SI Appendix Fig. S19). There is still an enhancement in $k_{obs,ST1}$ at 50°C, albeit less than that at 42°C, despite being closer to the T_m of ST1 . We modeled the thermostability of Pus7, and our model indicates that Pus7 is not stable at 50°C (SI Appendix Fig. S20), likely accounting for the decrease in activity observed at this temperature.

Consistent with our *in vitro* observations, we modeled the structures of 20 randomly selected mRNAs reported to be modified by Pus7 under heat shock at 30°C and 45°C (SI Appendix Fig. S20). Comparison of these models reveals that most of the targeted Us we modeled are in different, often less structured, contexts at 45°C than at 30°C (SI Appendix Fig. S13). We further tested this by measuring the degree of Ψ incorporation on nine full-length Pus7 substrates (ARG5,6, BET2, TEF5, $RTC3_{(U77)}$, $RTC3_{(U288)}$, $TEF2_{(U555)}$, $TEF2_{(U1104)}$, $U2snRNA_{(U35)}$, $U2snRNA_{(U56)}$) at different temperatures by CLAP (42). In these assays, we reacted 150 μg of total RNA purified from *pus7*Δ cells with 50 μM of purified Pus7 at 30°C and 37°C for 10 minutes. The extent of pseudouridylation was increased at 37°C on all but one of the 'heat shock' targets with more predicted secondary structure (TEF5, $RTC3_{(U77)}$, $RTC3_{(U288)}$, $TEF2_{(U555)}$, $TEF2_{(U1104)}$) (Fig. 6B). In contrast, the level of pseudourdinylation on targets previously observed in Ψ-mapping

studies at 30°C (ARG5,6, BET2, U2snRNA$_{(U35)}$) or under nutrient starvation (U2snRNA$_{(U56)}$) were either unaffected or decreased (Fig. 6B).

DISCUSSION

Our fundamental understanding of the structure and mechanism of pseudouridine synthases is largely built on foundational studies of bacterial enzymes that exclusively modify non-coding RNAs. The discovery of pseudouridine in eukaryotic mRNAs, coupled with the identification of heritable diseases caused by mutations to human Pus enzymes have ignited a renewed interest in eukaryotic pseudouridine synthases. While all Pus enzymes share a structurally conserved catalytic core, the eukaryotic enzymes possess additional insertions with unidentified functions. Pus7, a homolog of the bacterial pseudouridine synthase TruD, is among the enzymes that modify the largest number of mRNAs in eukaryotes (23, 24). We solved the crystal structure of *S. cerevisiae* Pus7, revealing the inherent flexibility in the form and position of Pus7 eukaryotic-specific insertions (Fig. 1B). The largest insertion (ID-I), is a positively charged domain connected to the Pus7 core by long flexible linkers. Normal mode analyses of our structure with elNémo and DynOmics Portal 1.0, along with the differing positions of ID-I observed in the two Pus7 structures, suggest that this domain is very flexible. ID-I appears capable of swinging away the core of the molecule, which may allow it to form contacts with RNA substrates distal from their modification sites. Our findings that the removal of ID-I ($Pus7_{\Delta IDI}$) reduces RNA binding and enhances substrate discrimination are consistent with such a model (Fig. 5, Table 1). Nonetheless, in contrast to the large effect of active site mutations on substrate catalysis (38- to 74,000-fold), the impacts of ID-I deletion both *in vivo* and *in vitro* are modest (2- to 4-fold reduction in Ψ incorporation) and indicate that the role of ID-I is more likely to fine-tune RNA substrate selection (Fig. 5 and SI Appendix Fig S13).

Pus7 is distinguished from other pseudouridine synthases by the apparent variety of substrates that it has been reported to modify. Our *in vitro* biochemical assays support this idea, as we find that Pus7 can bind and pseudouridylate UGUAR motifs in diverse sequence and structural contexts (Fig. 4). Our kinetic and CLAP data indicate that although Pus7 can modify uridines predicted to be in strong secondary structures, it is most active on Us in regions with less predicted structure (Figs. 4, 6 and SI Appendix Fig. S19). The relatively slow k_{obs} value we measured for $tRNA^{Asp}$, where the targeted uridine is base-paired, exemplifies this trend. These biochemical observations are in line with the predicted structural contexts of uridines targeted by Pus7 in yeast cells. The models we obtained using RNAStructure reveal a trend for Pus7 substrate selection in which modified UGUAR sequences are in less structured regions than unmodified UGUAR sequences (SI Appendix Fig. S16). These observations can help to partially rationalize which UGUAR sites Pus7 modifies in cells.

Under heat-shock, Pus7 modifies 15-fold more mRNAs (24). Consistent with this, we found that the ST1 RNA, predicted to contain a base-paired UGUAG sequence, is modified more efficiently at elevated temperatures, where the stability of the base-paired region of the molecule is significantly reduced (Fig. 6A, SI Appendix Fig. S19). Additionally, we measured the stoichiometry of Ψ-incorporation in full-length RNAs previously reported to be modified under either unstressed or heat-shock conditions (23, 24). We saw that heat-shock induced sites are more efficiently modified at higher-temperatures than Ψ-sites detected in unstressed cells (Fig. 6B). This is notable because Pus7 should be less stable (SI Appendix Fig. S20) and presumably less active

at elevated temperatures. Our data suggest that increased RNA dynamics are more important than having optimal enzyme activity on these substrates.

Although our data indicate that Pus7 more quickly modifies structurally unconstrained Us, we find that if left for long enough (2-10 minutes) Pus7 converts 100% of Us to Ψs in all of the model UGUAG-containing sequences we present it (Fig. 4). Similarly, we also observed that Pus1 is able to modify sequences not predicted to contain its preferred secondary structure if allowed to react for 30 minutes (SI Appendix Fig. S9). Since a significant portion of Us identified as Pus7 targets in sequencing studies are predicted to be in structured regions, a simplistic model in which Pus7 only interacts with single stranded uridines does not satisfactorily explain either our *in vitro* studies or the breadth of targets identified by Ψ-mapping in cells. Our findings suggest that instead of identifying motifs that Pus7 can modify, we need to address why this promiscuous enzyme does *not* modify every UGUAR sequence in cells. We propose that Pus7 rapidly samples RNA sequences and opportunistically selects substrates that contain an accessible (even if only transiently) UGUAR sequence motif (SI Appendix Fig S15). Such a mechanism is reminiscent of DNA glycosylases that use facilitated diffusion to quickly scan nonspecific sites in their search for damaged bases (47, 48).

The work that we present here suggests that Pus7 is a promiscuous enzyme, and that factors beyond inherent enzyme properties (e.g. enzyme localization, RNA structure, and competition with other RNA-binding proteins) significantly contribute to shaping Pus7 substrate selection (49, 50). This idea is exemplified by the observed re-localization of Pus7 to the cytoplasm, and subsequent increased substrate scope under heat shock (24). The importance of protein localization and cellular conditions to pseudouridine synthase target selection is unlikely to be unique to Pus7. Pus4 was recently reported relocalize to the cytoplasm and have increased activity towards its mRNA substrates when it is in a prion conformation (18). Collectively, these findings indicate that the environment of potential RNA substrates, which remodels in response to changing cellular and environmental conditions, plays a previously unrecognized role in determining the Ψ-modification landscape.

METHODS

Native and SeMet labeled *S. cerevisiae* Pus7 and Pus 1 proteins were expressed in *E. coli* BL21(DE3) cells and purified on sequential Ni(NTA) and Resource Q columns. Pus7 crystals were obtained by the sitting-drop vapor-diffusion. Diffraction data were collected at Advanced Photon Source and processed using XDS. Phaser was used to obtain a molecular replacement solution and REFAC5 was used for model refinement. The quality of the final Pus7 model was assessed with MolProbity. Wild-type and *pus7*Δ growth were assessed in YPD supplemented with either NaCl, $MgSO_4$, puromycin, cycloheximide, hygromycin B, MG132 and 1.5-3 mg/mL paromomycin, or YEP supplemented with either glucose, galactose or sucrose. All RNAs (SI AppendixTable S5) were prepared by run-off T7 transcription of DNA oligonucleotides, and labeled (5' fluorescein or ^3H) for binding, stopped-flow and single-turnover experiments. Pus7 substrate binding was evaluated by EMSA and stopped-flow and modification was monitored by ^3H-release. The stoichiometries of Ψ incorporation at specific sites in full-length mRNAs modifications were measured using CMC-RT and Ligation Assisted PCR analysis of Ψ modification (CLAP). Detailed procedures and reaction conditions for all experiments are provided in the SI Appendix Methods.

TABLES AND LEGENDS

Table 1

PUS7	k_{obs} (s^{-1})*	$K_{D,app1}$ (nM)†
WT	$9.9 \pm 1.0 \times 10^{-1}$	75 ± 15
K61A	$2.6 \pm 0.1 \times 10^{-2}$	180 ± 40
F67A	$4.6 \pm 0.2 \times 10^{-3}$	480 ± 50
E71A	$5.2 \pm 0.3 \times 10^{-3}$	210 ± 50
H161A	$6.9 \pm 0.9 \times 10^{-1}$§	170 ± 40
D256A	N.A.**	60 ± 15
N305A	$4.0 \pm 0.1 \times 10^{-4}$	230 ± 60
F307A	$1.3 \pm 0.1 \times 10^{-5}$	340 ± 170
F307Y	$2.6 \pm 0.1 \times 10^{-3}$	N.D.***
ΔID1	$3.8 \pm 0.6 \times 10^{-1}$§	160 ± 40

* k_{obs} values were determined by tritium release assays using 2 μM PUS7 and <100 nM CDC8 substrate. At least three replicate curves were collected for each mutant.

† $K_{D,app1}$ values were determined by electrophoretic mobility shift assays ($n \geq 2$) using PUS7 containing the catalytically inactive D256A mutation in addition to the mutation indicated in the first column.

§ k_{obs} value determined using 10 μM PUS7 and <100 nM CDC8 substrate.

** N.A. no activity

***N.D., not determined.

ACKNOWLEDGEMENTS

This work was funded by the University of Michigan start-up funds (K.S.K. and M.K.), Rackham Merit Fellowship (M.K.F.), National Institutes of Health awards R35 GM128836 (K.S.K.), R01 GM117141 (M.K.), T32 GM008597 (M.K.F.) and Research Corporation for Science Advancement Cottrell Scholar Award (K.S.K.). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). Biorender was used to create Figures 5A and 6.

REFERENCES

1. Jones JD, Monroe J, & Koutmou KS (2020) A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. *Wiley Interdiscip Rev RNA* 11(4):e1586.
2. Martinez NM*, et al.* (2020) Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect splicing. *bioRxiv*.
3. Levi O & Arava YS (2021) Pseudouridine-mediated translation control of mRNA by methionine aminoacyl tRNA synthetase. *Nucleic Acids Res* 49(1):432-443.
4. Nakamoto MA, Lovejoy AF, Cygan AM, & Boothroyd JC (2017) mRNA pseudouridylation affects RNA metabolism in the parasite Toxoplasma gondii. *RNA* 23(12):1834-1849.
5. Gutgsell N*, et al.* (2000) Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. *RNA* 6(12):1870-1881.
6. Del Campo M, Kaya Y, & Ofengand J (2001) Identification and site of action of the remaining four putative pseudouridine synthases in Escherichia coli. *RNA* 7:1603-1615.
7. Breslow DK*, et al.* (2008) A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* 5(8):711-718.
8. Urbonavicius J, Durand JM, & Bjork GR (2002) Three modifications in the D and T arms of tRNA influence translation in Escherichia coli and expression of virulence genes in Shigella flexneri. *J Bacteriol* 184(19):5348-5357.
9. Ishida K*, et al.* (2011) Pseudouridine at position 55 in tRNA controls the contents of other modified nucleotides for low-temperature adaptation in the extreme-thermophilic eubacterium Thermus thermophilus. *Nucleic Acids Res* 39(6):2304-2318.
10. Kinghorn SM, O'Byrne CP, Booth IR, & Stansfield I (2002) Physiological analysis of the role of truB in Escherichia coli: a role for tRNA modification in extreme temperature resistance. *Microbiology (Reading)* 148(Pt 11):3511-3520.
11. Sinha H*, et al.* (2008) Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics* 180(3):1661-1670.
12. de Brouwer APM*, et al.* (2018) Variants in PUS7 Cause Intellectual Disability with Speech Delay, Microcephaly, Short Stature, and Aggressive Behavior. *Am J Hum Genet* 103(6):1045-1052.
13. Zeharia A*, et al.* (2005) Mitochondrial myopathy, sideroblastic anemia, and lactic acidosis: an autosomal recessive syndrome in Persian Jews caused by a mutation in the PUS1 gene. *J Child Neurol* 20(5):449-452.
14. Bykhovskaya Y, Casas K, Mengesha E, Inbal A, & Fischel-Ghodsian N (2004) Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *Am J Hum Genet* 74(6):1303-1308.
15. Shaheen R*, et al.* (2016) A homozygous truncating mutation in PUS3 expands the role of tRNA modification in normal cognition. *Hum Genet* 135(7):707-713.
16. Keffer-Wilkes LC, Veerareddygari GR, & Kothe U (2016) RNA modification enzyme TruB is a tRNA chaperone. *Proc Natl Acad Sci U S A* 113(50):14306-14311.
17. Keffer-Wilkes LC, Soon EF, & Kothe U (2020) The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain. *Nucleic Acids Res* 48(14):7981-7990.
18. Garcia DM*, et al.* (2021) A prion accelerates proliferation at the expense of lifespan. *Elife* 10.
19. Penzo M, Guerrieri AN, Zacchini F, Trere D, & Montanaro L (2017) RNA Pseudouridylation in Physiology and Medicine: For Better and for Worse. *Genes (Basel)* 8(11).
20. Hamma T & Ferre-D'Amare AR (2006) Pseudouridine synthases. *Chem Biol* 13(11):1125-1135.
21. Rintala-Dempsey AC & Kothe U (2017) Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol* 14(9):1185-1196.

22. Boccaletto P, *et al.* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* 46(D1):D303-D307.
23. Carlile TM, *et al.* (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515(7525):143-146.
24. Schwartz S, *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159(1):148-162.
25. Safra M, Nir R, Farouq D, Vainberg Slutskin I, & Schwartz S (2017) TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res* 27(3):393-406.
26. Carlile TM, *et al.* (2019) mRNA structure determines modification by pseudouridine synthase 1. *Nat Chem Biol* 15(10):966-974.
27. Guzzi N, *et al.* (2018) Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. *Cell* 173(5):1204-1216 e1226.
28. Behm-Ansmant I, *et al.* (2003) The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA* 9(11):1371-1382.
29. Ma X, Zhao X, & Yu YT (2003) Pseudouridylation (Psi) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism. *EMBO J* 22(8):1889-1897.
30. Kaya Y & Ofengand J (2003) A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* 9(6):711-721.
31. Ericsson UB, Nordlund P, & Hallberg BM (2004) X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. *FEBS Lett* 565(1-3):59-64.
32. Hoang C & Ferre-D'Amare AR (2004) Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA* 10(7):1026-1033.
33. Kaya Y, Del Campo M, Ofengand J, & Malhotra A (2004) Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J Biol Chem* 279(18):18107-18110.
34. Mueller EG (2002) Chips off the old block. *Nat Struct Biol* 9(5):320-322.
35. Holm L (2020) Using Dali for Protein Structure Comparison. *Methods Mol Biol* 2112:29-42.
36. Yoshikawa K, *et al.* (2009) Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in Saccharomyces cerevisiae. *FEMS Yeast Res* 9(1):32-44.
37. Blanchet S, *et al.* (2018) Deciphering the reading of the genetic code by near-cognate tRNA. *Proc Natl Acad Sci U S A* 115(12):3018-3023.
38. Chou HJ, Donnard E, Gustafsson HT, Garber M, & Rando OJ (2017) Transcriptome-wide Analysis of Roles for tRNA Modifications in Translational Regulation. *Mol Cell* 68(5):978-992 e974.
39. Eyler DE, *et al.* (2019) Pseudouridylation of mRNA coding sequences alters translation. *Proc Natl Acad Sci U S A* 116(46):23068-23074.
40. Czekay DP, Schultz SK, & Kothe U (2021) Assaying the Molecular Determinants and Kinetics of RNA Pseudouridylation by H/ACA snoRNPs and Stand-Alone Pseudouridine Synthases. *Methods Mol Biol* 2298:357-378.
41. Veerareddygari GR, Singh SK, & Mueller EG (2016) The Pseudouridine Synthases Proceed through a Glycal Intermediate. *J Am Chem Soc* 138(25):7852-7855.
42. Zhang W, Eckwahl MJ, Zhou KI, & Pan T (2019) Sensitive and quantitative probing of pseudouridine modification in mRNA and long noncoding RNA. *RNA* 25(9):1218-1225.
43. Bellaousov S, Reuter JS, Seetin MG, & Mathews DH (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res* 41(Web Server issue):W471-474.
44. Wright JR, Keffer-Wilkes LC, Dobing SR, & Kothe U (2011) Pre-steady-state kinetic analysis of the three Escherichia coli pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis. *RNA* 17(12):2074-2084.
45. Huang L, Pookanjanatavip M, Gu X, & Santi DV (1998) A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry* 37(1):344-351.

46. Ramamurthy V, Swann SL, Spedaliere CJ, & Mueller EG (1999) Role of cysteine residues in pseudouridine synthases of different families. *Biochemistry* 38(40):13106-13111.
47. Taylor EL, Kesavan PM, Wolfe AE, & O'Brien PJ (2018) Distinguishing Specific and Nonspecific Complexes of Alkyladenine DNA Glycosylase. *Biochemistry* 57(30):4440-4454.
48. Esadze A & Stivers JT (2018) Facilitated Diffusion Mechanisms in DNA Base Excision Repair and Transcriptional Activation. *Chem Rev* 118(23):11298-11323.
49. Vaidyanathan PP, AlSadhan I, Merriman DK, Al-Hashimi HM, & Herschlag D (2017) Pseudouridine and N(6)-methyladenosine modifications weaken PUF protein/RNA interactions. *RNA* 23(5):611-618.
50. Gerber AP, Herschlag D, & Brown PO (2004) Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* 2(3):E79.

FIGURES LEGENDS

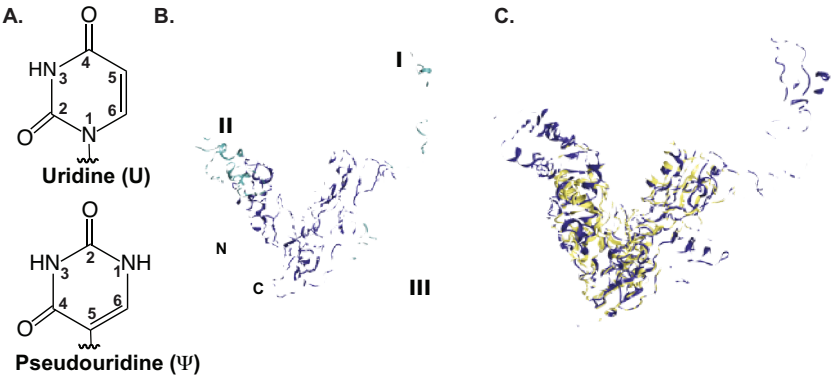


Figure 1. *S. cerevisiae* Pus7 structure.
(A) Uridine and pseudouridine. **(B)** X-ray structure of Pus7 at 3.2Å resolution (PDB 7MZV). The structurally conserved V-shaped enzyme core housing the PUS and TRUD domains (blue). The three eukaryotic specific insertions (green) are numbered I-III. **(C)** Superimposition of the *S. cerevisiae* Pus7 (blue) and *E. coli* TruD (yellow, PDB:1SB7) structures demonstrating the structural conservation of the enzyme's catalytic core.

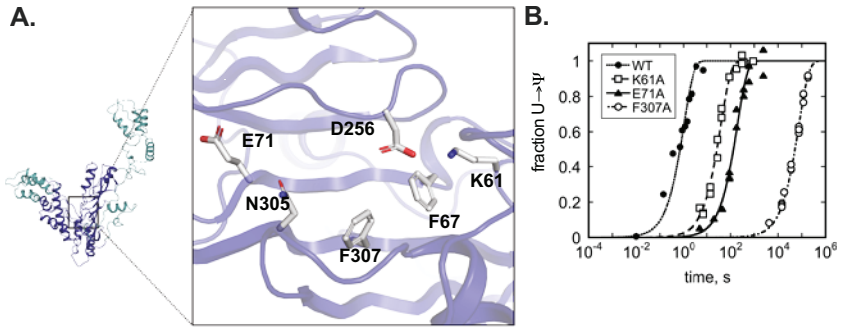


Figure 2. Pus7 active site residues enhance catalysis.
(A) Conserved Pus7 active site residues investigated in this study. **(B)** Time-courses for Ψ incorporation into a CDC8 mRNA by saturating concentrations of (●) wild-type Pus7 and Pus7 active site mutants ((□) K61A, (▲) E71A, (o) F307A). The single-turnover rate constants (k_{obs}) for alanine substitutions of all of the residues displayed in **(A)** are reported in Table 1.

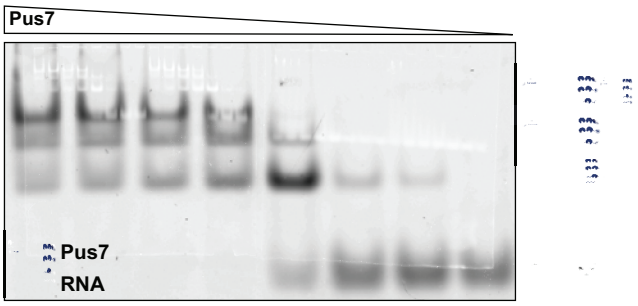


Figure 3. Multiple Pus7 proteins bind to CDC8 RNA.
The association of increasing concentrations of catalytically inactive D256A Pus7 with limiting amounts of 5'-fluorescein labeled CDC8 visualized on a non-denaturing agarose gel. Increased concentrations of D256A resulted in multiple binding events.

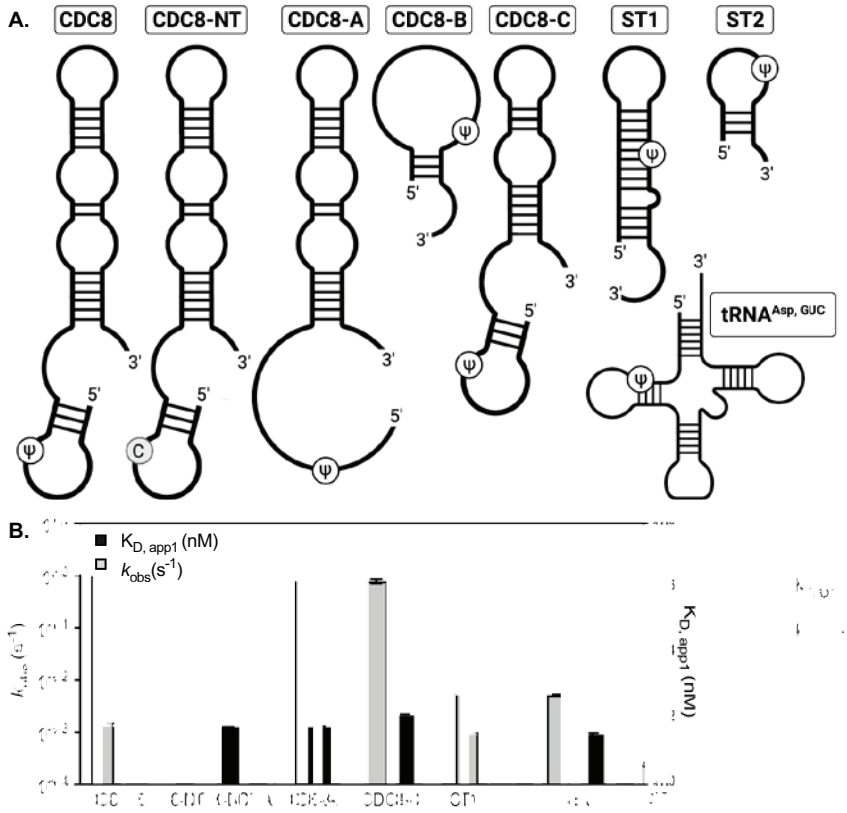


Figure 4. Pus7 can bind and modify a variety of RNA substrates.
(A) Secondary structures of the RNAs investigated in this study. The substrate sequences are available in SI AppendixTable S5. **(B)** $K_{D,app1}$ (right y-axis, black bars) and k_{obs} (left y-axis, gray bars) values for Pus7 binding and modifying the substrates displayed in **(A).** The $K_{D,app1}$ displayed for ST2 (*) is a lower limit for this value (SI AppendixFigure S9).

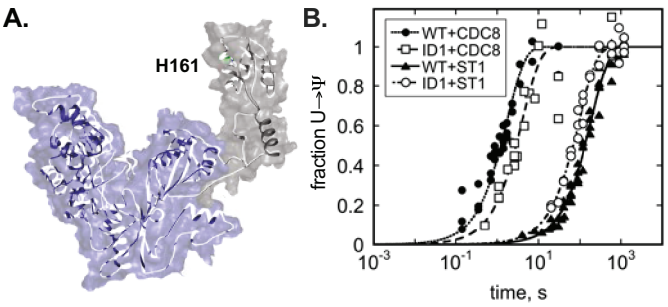


Figure 5. Insertion domain I enhances Pus7 selectivity for CDC8 over ST1.
(A) Crystal structure of Pus7 with ID-I shown in gray. The Pus7$_{\Delta ID1}$ protein (blue) lacks ID-I(gray). **(B)** Time courses of Ψ-incorporation into CDC8 and ST1 by wild-type Pus7 (●- CDC8, ▲ -ST2) and Pus7$_{\Delta ID1}$ (□- CDC8, o – ST1). The single-turnover k_{obs} values for these reactions are reported in SI AppendixTable S4. ID-I enhances the ability of Pus to discriminate between CDC and ST2; wild-type $k_{obs,CDC8} / k_{obs,ST2} = 178$, and Pus7$_{\Delta ID1}$ $k_{obs,CDC8} / k_{obs,ST2} = 40$.

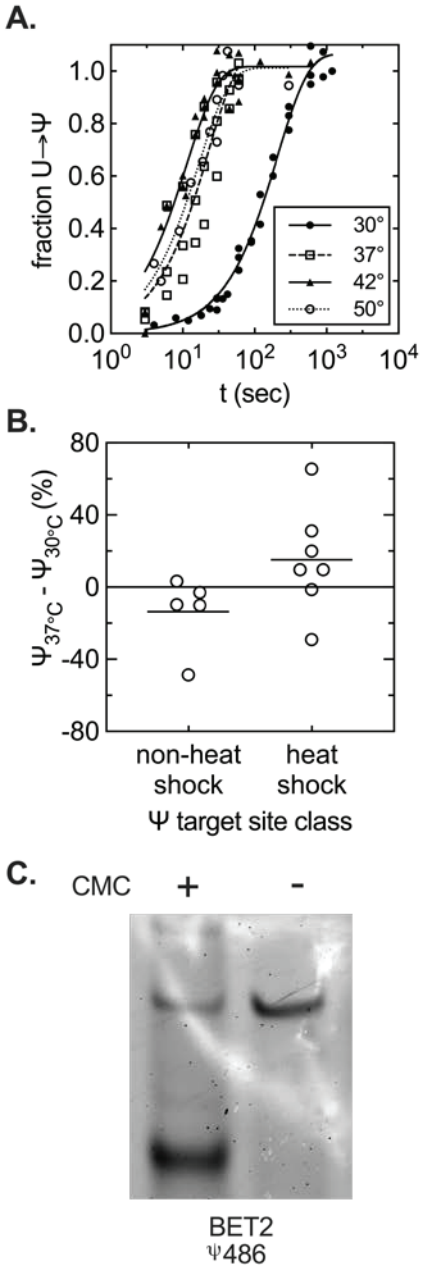


Figure 6. Pus7 is more active at elevated temperatures on substrates with UGUAR sequences predicted to be in secondary structures. (A) Time courses of Ψ-incorporation into ST1 by wild-type Pus7 at varying temperatures (●- 30ºC, □- 37ºC, ▲ - 42ºC, o- 50ºC). The single-turnover k_{obs} values for these reactions are reported in SI Appendix Fig. S19. **(B)** Differences in the stoichiometry of Ψ-incorporation at 30ºC and 37ºC in full-length RNA substrates measured by CLAP. The level of Ψ-addition is generally enhanced at sites that were only detectable under heat shock in Schwartz, et al. (1). **(C)** Representative CLAP gel of BET2 pseudouridylated target site from total RNA purified from BY4741 yeast Δpus7::kanMX. Black arrow denotes the truncated, pseudouridylated product. The upper band is the unmodified, full-length product.

SUPPLEMENTAL APPENDIX EXTENDED METHODS

Expression and purification of wild-type, mutant and truncated Pus7 proteins.

Saccharomyces cerevisiae wild-type and truncated ($Pus7_{\Delta N34C9}$, $Pus7_{\Delta IDI}$) Pus7 protein encoding DNA-sequences were ordered from GeneArt. Ligation independent cloning was used to incorporate these sequences into a pMCSG7 vector containing an N-terminal His_6-tag and TEV cleavage site. Single and double mutants were incorporated into the Pus7 sequence by QuikChange® site-directed mutagenesis (Stratagene) using appropriate primers (IDT) (SI Methods). Sequences were confirmed by Sanger DNA sequencing (UMich sequencing core). All proteins were expressed in BL21(DE3)-P-LysS *E. coli* cells grown in 1L Terrific Broth, 100 µg/mL ampicillin at 37°C and 250 RPM. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.15 mM when cells reached OD_{600} of ~0.6. Following induction, cells were grown for 18 hours at 20°C and harvested by centrifugation at 5,000 RPM for 30 minutes. Pus7 proteins were purified on a Ni^{2+} Hi-Trap column (GE healthcare), the His-tag was removed by TEV protease treatment followed by a second Ni^{2+} Hi-Trap column. The protein was further purified by anion exchange chromatography on a 5 mL Resource™ Q column (GE Healthcare), and size exclusion chromatography using a Superdex 200 column (GE Healthcare). Purified protein was either concentrated and stored at -80° C or used immediately for crystallization.

Selenomethionine Expression

pMCSG7-yPus7 was expressed in BL21(DE3) cells grown in Terrific Broth media (4% glycerol), 100 µg/mL ampicillin at 37°C overnight. The cells were pelleted resuspended in 1.1L of selenomethionine minimal media, supplemented with 50 µg/mL L-selenomethionine, and 100 mL of freshly prepared, and sterile filtered nutrient solution 20% (w/v) glucose, 0.3% (w/v) $MgSO_4$, 0.1mg/mL $Fe(II)(SO_4)_3$, 0.1 mg/mL Thiamine, adjust to pH 7.4, sterile. The cells were then grown at 37°C and 250 RPM until OD_{600} of 0.6. The cells were induced with IPTG to a final concentration of 0.2 mM and grown for 18 hours at 20°C before harvesting by centrifugation.

Crystallization

Unlabeled and SeMet derivatized Pus7 was concentrated 10 mg/mL in 50 mM TRIS, pH 7.5, 50 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP). Crystals of Pus7 were obtained by the sitting-drop vapor-diffusion method at 20°C by mixing 0.5uL of protein solution (10mg/mL) with 0.5uL of the reservoir solution which contained 2 M ammonium sulfate, 10 mM nickel (II) chloride, 100 mM TRIS pH 8.5. The crystals were then cryoprotected in a solution of 15% glycerol, 1.7 M ammonium sulfate, 0.85 mM nickel (II) chloride, 85 mM TRIS pH 8.5 before being flash cooled in liquid-nitrogen.

Crystal data processing

Diffraction data were collected at 100 K and at the Se edge on LS-CAT 21-ID-D at Advanced Photon Source, Argonne National Laboratory using a DECTRIS EIGER 9M. Three data sets were collected from two crystals, and all were separately processed with XDS to 3.2 Å resolution were Friedel pairs were treated as equal. Reflections from a total of 1500 selected frames (first 500 from 2 datasets and first 400 from the third) were merged and scaled with Aimless (cite) and the

resulting reflection file was used for subsequent refinements of our Pus7 model. The data were indexed to space group C222 (unit-cell parameters *a* = 117.9, *b* = 171.8, *c* = 105.3 Å) with 1 molecule in the asymmetric unit (Matthew's coefficient V_M = 3.46 Å^3 Da^{-1}, 64.5% solvent content). 500 frames from a single data set were processed anomalous (Friedel pairs were not treated as equal) with XDS to 3.2 Å and the resulting reflection file was used for the SAD phasing.

Crystal structure solution

Initial structure solutions were obtained by molecular replacement using the human Pus7 (PDB:5KKP) as a search model and initial phases were calculated using Phaser (1). However, we were unable to obtain a structure solution for insertion domain one, which necessitated the growth of Se-Met Pus7 crystals. AutoSol (2) was used to identify selenium sites and calculate density-modified 3.3 Å experimental maps based on a single-wavelength single-wavelength anomalous dispersion (SAD) data set from SeMet Pus7 (the experimentally determined SeMet f' and f" values that were used were -7.4 and 5.0 respectively). Specifically, 16 selenium sites were located and used for SAD phasing, using phenix.hyss. Subsequently, Phaser was used to calculate the experimental phases, followed by density modification with RESOLVE (figure of merit 0.36 before and 0.78 after density modification). The experimental density map showed clear features of the protein backbone and well-defined side chains. RESOLVE traced and automatically built 389 residues and their side chains in the experimental electron density. The final experimental model was in really good agreement with our original MR derived model but also provided us with a partial model of ID-1. The partial model of ID-1 included residues 129 to 148, a region of ID-1 that packs against the core of an adjacent monomer and includes the only SeMet present in ID1. The electron density corresponding to the insertion domain is overall poor and of rather low resolution, as also reflected in the very high average B-factors (165.02) as compared to the average B-factors (117.44) for the rest of the protein (Figure S2E). Ultimately, using SAD phasing, in combination with our MR model, we were able to obtain a structure solution for the insertion domain, completing our structure model. An overlay of the final Pus7 model with all 16 experimentally determined selenium heavy atoms is shown in Fig. S2. The structural model was refined with REFMAC5 as part of the CCP4I2 package (3) using isotropic individual *B*-factors with maximum-likelihood targets where the Babinet model for bulk-solvent scaling was utilized. Refinement was followed by model building and modification with Coot (4). We performed several iterative rounds of refinement followed by model building and modification. All crystallographic information as well as refinement statistics are provided in Table 1. The geometric quality of the model and its agreement with the structure factors were assessed with MolProbity (5). Figures displaying crystal structures were generated by PyMOL(6).

Preparation of 5'-fluorescein labeled RNA substrates.

RNA was prepared via *in vitro* transcription from DNA oligonucleotide templates ordered from Integrated DNA Technologies (IDT) and transcribed by recombinant T7 RNA polymerase (7). Transcription reactions were carried out in 50 mM Tris-HCl pH 8.0, 4 mM $MgCl_2$, 1 mM spermidine, 5 mM DTT, 4 mM ATP, 4 mM CTP, 4 mM UTP, 1 mM GTP, 4 mM guanosine-5'-O-monophosphorothiolate (GMPS), 350 µg/mL purified T7 RNA polymerase, 12.5 µM purified DNA template containing T7 promoter and 4 U/µl SUPERaseIn. After stopping the transcription by the addition of 50 mM EDTA and 500 mM NaCl, the RNA was washed with degassed TE pH 7.2 three

times using Amicon spin column (10 kDa MWCO). The washed RNA (~250 µl) was incubated with 20 µl 45 mM fluorescein overnight at 37°C to label the 5'end. All following steps were carried out in the dark. The reaction was stopped by addition of an equal volume of 2X loading dye (0.05% Bromophenol Blue, 0.05% Xylene Cyanol dye, 50% m/v urea, 0.1 M EDTA) and run on a 12% urea-polyacrylamide gel. The RNA was eluted via crush-and-soak method into buffer (TE, 0.1% SDS, and 0.5 M NaCl) overnight at 4°C. The elution products were subsequently filtered, washed, and concentrated using degassed TE and an Amicon spin column (10 kDa MWCO). The RNA was then ethanol precipitated at -20°C for 12 hours. The resulting pellet was resuspended in 20 µl of RNase free H_2O. The concentration of the total and labeled RNA were measured photometrically using A260 and A494 respectively, using a Nanodrop spectrophotometer. Select Fl-labeled substrates were also purchased from Dharmacon.

Electrophoretic mobility shift assays (EMSAs)

For gel-shift assays, serially diluted protein (0-2000 µM) was incubated with 10 nM of 5'-fluorescein labeled RNA in 10 µL reaction volumes for ≥ 5 min at 25°C in a binding buffer containing 100 mM NH_4OAc, 100 mM Tris, pH 8.0, 5 mM $MgCl_2$, 2 mM DTT, and 6% (w/v) sucrose. An aliquot of each reaction (5 µL) was loaded on a preequilibrated, native 6% polyacrylamide (37.5:1) gel in 1xTBE. Gels were electrophoresed at 30V for ~4h at 4°C. When fluorescently labeled RNA substrates were used, electrophoresis was performed in the dark. Gels were then rinsed in 1xTBE and imaged on an Amersham Typhoon Biomolecular Imager (GE Healthcare). If unlabeled RNA was used, the gel was stained with SYBR® Gold Nucleic Acid Gel Stain (Invitrogen) in 1xTBE for ≥ 30 min in the dark before imaging on the Typhoon. Band intensities were quantified using ImageQuant (Cytiva) and the percentage of RNA bound calculated using Equation 1:

$$RNA_{bound}(\%) = 100 \times \frac{{E_T}^{n_h}}{{K_{D,app}}^{n_h} + {E_T}^{n_h}}$$

Binding data were fit using equations derived from the binding models shown in Figure S10. In general, simpler models were tried first, and if systematic errors remained in the fit, more complex models were used to fit the data. The simplest model used was a Hill curve, Equation 2:

$$RNA_{bound}(\%) = 100 \times \frac{{E_T}^{n_h}}{{K_{D,app}}^{n_h} + {E_T}^{n_h}}$$

In this model, $K_{D,app}$ is the apparent K_D for binding of Pus7 to one of the many sites on a given RNA; $K_{D,app}{}^{n}{}_{h}$ is the concentration of Pus7 at which 50% of available sites are bound. When systematic errors remained in the fit, a more complex model was used in which Pus7 bound first

to a single specific site on the RNA, followed by the binding of multiple Pus7 moieties to multiple nonspecific sites on the same RNA (Figure S10B). These data were analyzed using Equation 3:

$$RNA_{bound}(\%) = 100 \times \frac{\left(E_T / K_{D,app1}\right) \times \left(1 + \left(E_T / K_{D,app2}\right)^{n_h}\right)}{1 + \left(E_T / K_{D,app1}\right) \times \left(1 + \left(E_T / K_{D,app2}\right)^{n_h}\right)}$$

Neither of these models are theoretically correct, in particular because there is no evidence for cooperative binding of Pus7 to RNA. A theoretically correct binding model would need to account for random binding of Pus7 to all of the possible binding sites on a given RNA. Each RNA has many binding sites, which are not all equivalent because of differences in sequence and structure, and the binding sites can interact with one another negatively (via steric occlusion, for example) and positively (e.g. binding of Pus7 at one site changes structure at a second site, increasing binding affinity). Our experimental methods do not provide enough information to develop such a model. The simplified models we use to analyze the data are therefore the best available tool, and allow for quantitative comparison of differences in binding that are identifiable via visual inspection of EMSA gel images.

Single-turnover pseudouridinylation assays

RNA substrates containing 5,6-[^{3}H]-uridine were prepared by *in vitro* transcription (7) and denaturing gel purification. Reaction buffer was as described for the EMSA experiments. RNA substrates were folded in 1X reaction buffer by heating to 60°C for 5 minutes, followed by a 30 minute incubation at 30°C (8). Indicated concentrations of protein were mixed with the smallest detectable amount of substrate (~3,000 cpm per uridine in each timepoint, which allows reliable detection of tritium release above 5% turnover). At each timepoint an aliquot of reaction mix (containing ~3,000 cpm/U) was quenched in 1,250 µL 0.1 M HCl (final) containing 250 µg Norit-A. Quenched timepoints were mixed, centrifuged at 21,000 x *g* for 5 minutes, and 1000 µL of supernatant was transferred to a new tube containing 250 µL of 0.1 M HCl with 250 µg Norit-A. Mixing and centrifugation were repeated, and 1000 µL of the supernatant was filtered through glass wool in a 1 mL pipet tip to remove residual charcoal. Aliquots of the filtrate (500 µL) were removed for liquid scintillation counting in a Beckman LSC-6500. For each reaction mix, input controls were prepared by passing an aliquot of reaction through the same process using 0.1 M HCl without the Norit-A. Counts observed in the input sample are used to calculate cpm/uridine, allowing calculation of the amount of Ψ produced at each timepoint. Background counts were determined by processing an RNA only reaction aliquot through the sample pipeline; these counts were routinely equivalent to background in our instrument (~30 cpm). Fraction of target U converted to Ψ data were fitted using Equation 4:

$$U > \Psi\ (fraction) = 1 - e^{-k_{obs} \times t}$$

Stop Flow Assays: Pus7/ D256A binding with fluorescently labeled mRNA

D256A Pus7 and 5'-fluorescein labeled CDC8 were generated and purified as described as above. Kinetic binding experiments were performed using the Kintek SF-300x stop-flow apparatus. Fluorescently labeled mRNA (5 nM final concentration) was mixed with D256A at

varied concentrations (20 nM – 750 nM final). Binding experiments were performed at room temperature in same buffer used in the EMSA experiments over the span of 1-1.5 seconds. Lower concentrations of Pus7/D256A (0-100nM) displayed monophasic behavior and were fit with a single exponential equation: $A_1e^{-k1t} + c$ to obtain a k_{obs1}. Higher concentrations displayed biphasic behavior and therefore were fit with a double exponential equation: $A_1e^{-k1t} + A_2e^{-k2t} + c$ to obtain k_{obs1} and k_{obs2}. The k_{obs1} values from both fits were then plotted against the concertation of D256A PUS7 mutant, displaying a linear relationship. The y-intercept gave a k_{off} of approximately 35 s^{-1} and the slope gave a k_{on} of ~4.3 x 10^8 $M^{-1}s^{-1}$. The K_D For D256A binding CDC8 was obtained using Equation 5: $K_D = k_{off}/k_{on}$.

Wild-type and *pus7*Δ growth assessment

Wild-type and *pus7*Δ yeast cells were inoculated into 3 mL YPD media and grown overnight. Then, they were diluted to OD_{600}=1 as a starting point, and 7 ml of 10-fold serial dilutions were spotted on fresh YPD agar plates supplemented with 0.75-1.0 M NaCl, 250 mM $MgSO_4$, 200 mM puromycin, 100 ng/mL cycloheximide, 25-50 mg/mL hygromycin B, 50 mM MG132 and 1.5-3 mg/mL paromomycin. Growth of the cells were also tested in the presence of different carbon sources including 2% glucose, 2% sucrose and 2% galactose in YEP agar media (1% yeast extract and 2% peptone). The plates were incubated for 2-3 days at 30°C unless otherwise indicated.

Phylogenetic tree generation

Annotated TruD/Pus7 sequences (>400 total sequences) from GenBank (NCBI) were aligned using ClustalW. Then, a representative 44 amino acid sequence was used for further analysis. Evolutionary analyses were conducted using MEGAX tool (9). The phylogenetic tree was generated using the Maximum Likelihood method (10). The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the taxa analyzed (11). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.3722)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.89% sites). There were a total of 1344 positions in the final dataset.

Ribosome profiling data analysis

Raw ribosome profiling sequencing data from two studies (12, 13) were downloaded and processed using the procedures described below. Briefly, adapter contaminations and low-quality reads were filtered out from the raw reads using the Cutadapt tool (14) like as previously described (15). Subsequently rRNA and tRNA contaminations were removed by aligning reads to the non-coding RNA (ncRNA) sequences of *S. cerevisiae* using Bowtie2 (16). Next, the remaining unaligned sequences were aligned against the transcriptome (coding RNA) of *S. cerevisiae* (R64-1-1 genome built) using TopHat2 (17). After that perfect match alignments were extracted from the TopHat output. For further downstream analysis, 3'- and 5'-end P-site offset values were

determined using riboWaltz (18). These P-site offset values are required to identify where ribosomes are located on each ribosome protected footprints (RPFs). After P-site offset calculation, actively translating ribosomes that represent trinucleotide periodicity were identified. Then the number of mapped RPFs was counted for each codon position within a gene using Samtools (19).

Modeling of thermal stability of PUS7

Using the established relationship between a protein's stability and its heat capacity (ΔC_p), stability (ΔG) chain length can be reasonably modeled as a function of chain length (N) and temperature (T) (26-29). Pus 7's stability curve was modeled as a function of N and T using previously published model seen in, Equation 6 (27, 29-31).

$$\Delta G(N,T) = \Delta H_R + \Delta C_p(T - T_R) - T\Delta S_R - T\Delta C_p \; ln \; ln \left(\frac{T}{T_R}\right) \quad (6)$$

Where enthalpy (ΔH_R) and entropy (ΔS_R) are calculated at a reference temperature and (ΔC_p) is the heat capacity of a protein, T_R is the reference temperature of 373 K for both ΔH_R and ΔS_R. These previous studies took advantage of the correlation between a protein's thermodynamic parameter and chain length to derive linear equations from experimental measurements collected into databases. The linear equations can be expressed as equations 7, 8, 9, (27, 29)

$$\Delta H_R = m_h \cdot N + b_h \quad (7)$$

$$\Delta S_R = m_s \cdot N + b_s \quad (8)$$

$$\Delta C_p = m_c \cdot N + b_c \quad (9)$$

Where m_h and b_h are the slope and intercept of ΔH_R, m_s and b_s are the slope and intercept of ΔS_R and m_c and b_c are the slope and intercept of ΔC_p when these parameters are plotted as a function of *N*. Equations 7, 8, 9 can be inserted into Equation 6, in order to get stability as a function of N and T as seen in Equation 10.

$$\Delta G(N, T) = \quad (10)$$
$$(m_h N + b_h) + (m_c N + b_c)(T - T_R) - T(m_s N + b_s) - T(m_c N + b_c) \; ln\left(\frac{T}{T_R}\right)$$

Detection and quantification of pseudouridylation: CLAP assay

The CLAP assay was adapted from Zhang, 2019 (25).

Pseudouridylation of total RNA or *in vitro* transcribed CDC8

Briefly, 150 μg of total RNA purified from BY4741 yeast *Δpus7::kanMX* was mixed with 50 μM Pus7-WT or Pus7-ΔID1 and incubated for 10 minutes at 30°C or 37°C in 1X pseudouridinylation buffer (100 mM TRIS-HCl pH 8.0, 100 mM NH_4OAc, 2 mM DTT, 5 mM $MgCl_2$) to modify the RNA. The reaction was stopped by adding 1/10th volume of 3 M NaOAc pH 5.2, followed by two phenol:chloroform (1:1) extractions with saturated acid phenol, and a final chloroform extraction to isolate the RNA. The RNA was then precipitated by adding an equal volume of 100% EtOH and 1 μL of GlycoBlue (Thermo Fisher, AM9515) and incubated at -20°C for 3 hours.

CMC treatment

RNA was resuspended in 41.5 μL of BEU buffer (50 mM Bicine pH 8.3, 4 mM EDTA, 7 M Urea). For CMC treated samples, 8.5 μL of freshly prepared 1 M CMC dissolved in BEU buffer was added, for a final concentration of 170 mM CMC. For CMC non-treated samples, 8.5 μL of BEU buffer was added, for a final reaction volume of 50 μL. Samples were incubated at 37°C for 20 minutes. The reaction was stopped by adding 100 μL of Stop Buffer (300 mM NaOAc pH 5.2, 0.1 mM EDTA) for a final volume of 150 μL. Excess CMC was removed by two sequential ethanol precipitations. Briefly, 700 μL 100% EtOH, and 1 μL GlycoBlue were added to the reaction before incubating 3 hours at -20°C. Sample was spun down for 30 min, 15kRPM, at 4°C before removing the supernatant, and washing the pellet by adding 500 uL of 70% EtOH, and spin for 5 min at 15kRPM. Remove supernatant and allow pellet to dry. Resuspend the RNA pellet in 100 μL of Stop Buffer and repeat ethanol precipitation and wash.

Alkali Treatment

Resuspend the pellet in 40 μL of 50 mM Na_2CO_3 pH 10.4 (pH taken at 37°C, temperature of incubation) and incubate for 3 hours at 37°C. Precipitate RNA via ethanol precipitation, as described above, with an additional 70% ethanol wash. Let pellet air dry. Resuspend the pellet in 20 μL sterile water and determine concentration by nano-drop.

RNA 5' Phosphorylation

To 6 μg RNA in 6.5 μL, add 1 μL 10X T4 PNK reaction buffer (NEB B0201S), 1 μL of 1 mM ATP, 0.5 μL 20 U/μL SUPERase•In RNase Inhibitor (Thermo Fisher AM2694), and 1 μL 10 U/μL T4 PNK (NEB M0201L) for a final volume of 10 μL. Incubate at 37°C for 30 minutes.

Blocker Ligation

To the reaction above, add 1 μL 10X T4 RNA Ligase reaction buffer (NEB B0216L), 1 μL of 100 μM 5' RNA blocker oligo (IDT /5AmMC6/rArCrCrCrA), 1 μL of 1 mM ATP, 1 μL 20 U/μL SUPERase•In RNase Inhibitor (Thermo Fisher AM2694), 3 μL DMSO, 2 μL sterile water and 1 μL 10 U/μL T4 RNA Ligase I (NEB M0204L) for a final volume of 20 μL. Incubate at 16°C for 16 h. Stop ligation reaction by adding 1.2 μL 200 mM EDTA.

Reverse transcription

For reverse transcription, the RT primer was first annealed by taking 3 μL of ligation mixture, adding 1 μL of 10 X annealing buffer (250 mM Tris-HCl, 480 mM KCl, pH 7.4) and 1 μL of 0.5 μM target specific reverse transcription primer (IDT). Samples were heated to 95°C for 3 minutes and slowly cooled to 37°C at a rate of -0.01°C/s (~15 min). To annealed sample, add 5 μL of 2 X AMV reverse transcription reaction mixture (1.2 U/μL AMV RT (NEB M0277L), 2X AMV RT buffer, and 1 mM of each dNTP) for a final concentration of 0.6 U/μL AMV RT, 1X AMV RT buffer, and 0.5 mM of each dNTP. Incubate at 42°C for one hour. Inactivate AMV RT by heating to 85 °C for 5 min before placing on ice. To digest RNA, add 1 μL of 5 U/μL RNaseH and incubate at 37°C for 20 minutes. Inactivate RNaseH by heating to 85 °C for 5 min and before placing reaction on ice. Add 1 μL of splint/adaptor oligo mix (1.5 μM adaptor oligo, 1.5 μM splint oligo) and incubate mixture at 75°C for 3 minutes followed 3 minutes at room temperature to anneal the splint/adaptor.

Add 4 μL of 4x ligation mixture (40 U/μL T4 DNA ligase, 4X T4 DNA ligase buffer, and 50% DMSO) for a final concentration of (10 U/μL T4 DNA ligase, 1X T4 DNA ligase buffer, and 12.5% DMSO). Incubate at 16°C for 16 h. Heat reaction to 65°C for 10 min to deactivate T4 DNA ligase, place immediately on ice.

PCR

Use 2 μL of reaction above, mix with 3.5 μL of 5 μM forward primer and 3.5 μL of 5 μM reverse primer (or reverse transcription primer). Add components for Q5 DNA polymerase reaction to a final volume of 35 μL and final concentration of 1X Q5 reaction buffer, 1X Q5 GC enhancer, 200 μM of each dNTP, 0.5 μM of forward and reverse primers, and 0.2 U/μL Q5 high fidelity DNA polymerase (NEB M0491L). Perform 35 cycles of PCR at requisite annealing temperatures for each site. 5 μL of PCR reaction was mixed with 1 μL of 6X TriTrack DNA loading dye and loaded on to a native 10% acrylamide (29:1) gel in 1X TBE pre-run at 10V/cm for 1 hour. Gel ran 3 hours at 10V/cm before being stained in 1X SYBR gold nucleic acid gel stain in 1X TBE for ~10 minutes. Gels were imaged on Amersham Typhoon imager and quantified using ImageQuant.

CLAP Primers

ARG5,6_RT	CCCATAGCAAGATTAATATTT
ARG5,6_FWD	TAGTTATTGGTGGTTTCA
ARG5,6_REV	TGCAGACATTGAGTAGC
ARG5,6_ADAPT	pCCATGTGAAACCACCAATAACTA
ARG5,6_SPLINT	TTTCACATGGAGTTGTTTGC/3SpC3/
BET2_RT	GCTTGAGCTGCATGGGATTCA
BET2_FWD	ACTATCAATTTTGGGTGAATTAA
BET2_REV	GCATTAGGACATAATCCAAAG
BET2_ADAPT	pCCATGTTAATTCACCCAAAATTGATAGT
BET2_SPLINT	ATTAACATGGAGACTTTGTA/3SpC3/
U2snRNA_RT	TATTATTTTGGGTGCCAAAAA
U2snRNA_56_FWD	CCTTTTGGCTTAGATCAA
U2snRNA_REV	ATGTGTATTGTAACAAATTAAAAGG
U2snRNA_56_ADAPT	pCCATGTTGATCTAAGCCAAAAGG
U2snRNA_56_SPLINT	ATCAACATGGAACAACTGAA/3SpC3/
U2snRNA_35_FWD	ACGAATCTCTTTGCCTTT
U2snRNA_35_ADAPT	pCCATGAAAGGCAAAGAGATTCGT
U2snRNA_35_SPLINT	CCTTTCATGGAGTATCTGTT/3SpC3/
CDC8_RT	ATATGCGTACTCAAAACAGGC
CDC8_FWD	GCTATTGGATAAAGAGATAAGGA
CDC8_REV	TCAACGATTTGCCAAATAAGC
CDC8_ADAPT	pCCATGTCCTTATCTCTTTATCCAATAGC
CDC8_SPLINT	AAGGACATGGAGACGTTACT/3SpC3/

EFB1/TEF5_81_RT	GTTGAACCATCTGGAGAATTC
EFB1/TEF5_81_FWD	GAAACAATTAAACGCTTCTTT
EFB1/TEF5_81_REV	TGGGTAAGCAGATTGGAAA
EFB1/TEF5_81_ADAPT	pCCATGAAAGAAGCGTTTAATTGTTTC
EFB1/TEF5_81_SPLINT	TCTTTCATGGACTGCTGTTT/3SpC3/
RTC3_77_RT	TCCTGAGGAGTGAAAACTTCG
RTC3_77_FWD	GGTGAAAATACAGATTTGATTG
RTC3_77_REV	AAGAGTTCGACAACTTCAGAT
RTC3_77_ADAPT	pCCATGCAATCAAATCTGTATTTTCACC
RTC3_77_SPLINT	GATTGCATGGAGACGAATAT/3SpC3/
RTC3_288_RT/REV	TCAATTGTAGGCTTTGGTTC
RTC3_288_FWD	GTTATCGATTTGATATTGAGAAA
RTC3_288_ADAPT	pCCATGTTTCTCAATATCAAATCGATAAC
RTC3_288_SPLINT	AGAAACCATGGAGTCTCAAAA/3SpC3/
TEF2_555_RT	GGACTTCAAGAACTTTGGATG
TEF2_555_FWD	GAAACCTCCAACTTTATCAA
TEF2_555_REV	GGTGGTAGCTTCAATCATGTT
TEF2_555_ADAPT	pCCATGTTGATAAAGTTGGAGGTTTC
TEF2_555_SPLINT	ATCAACATGGGTTCCATTCG/3SpC3/
TEF2_1104_RT	ACCCTTGTACCATGGAGCGTT
TEF2_1104_FWD	TTACTCTCCAGTTTTGGA
TEF2_1104_REV	GTCTTCCAACTTCTTACCAGA
TEF2_1104_ADAPT	pCCATGTCCAAAACTGGAGAGTAA
TEF2_1104_SPLINT	TTGGACATGGAGATTCGACG/3SpC3/

SUPPLEMENTAL APPENDIX FIGURES

Figure S1: Phylogenetic relations in TruD and Pus7 family. This tree shows the relation of Pus7 family proteins in different species. It also represents the relation between Pus7 family with TruD family proteins.

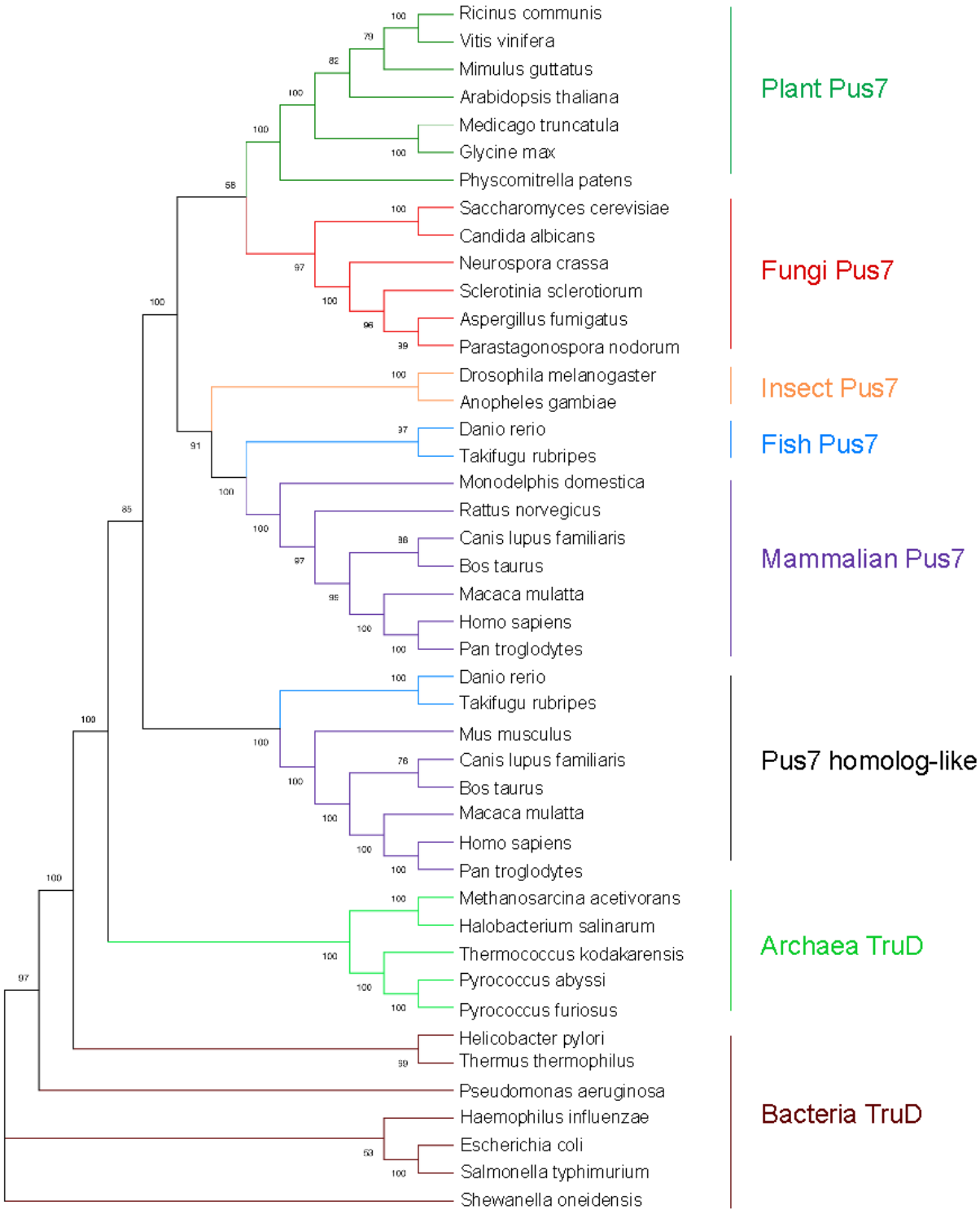
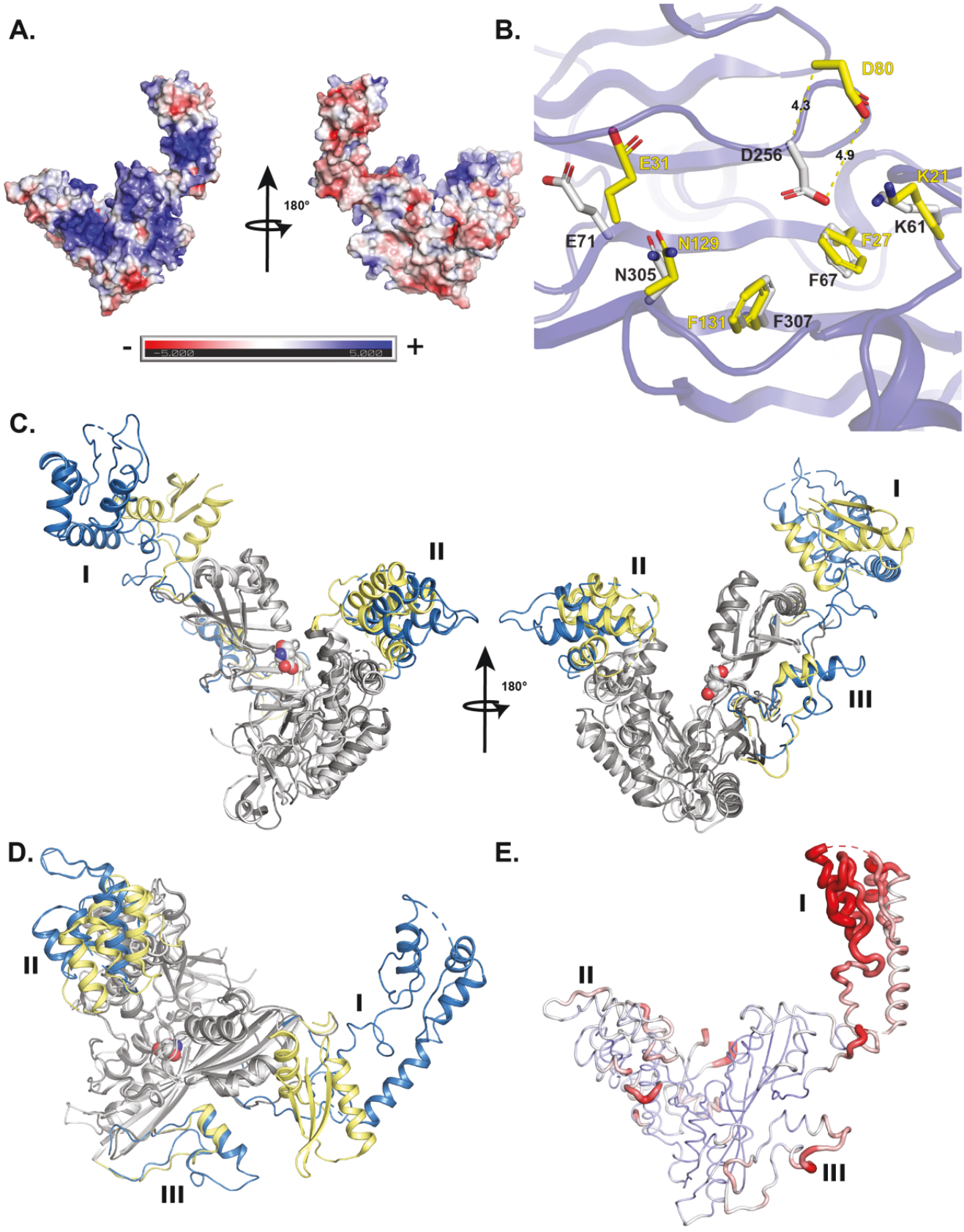
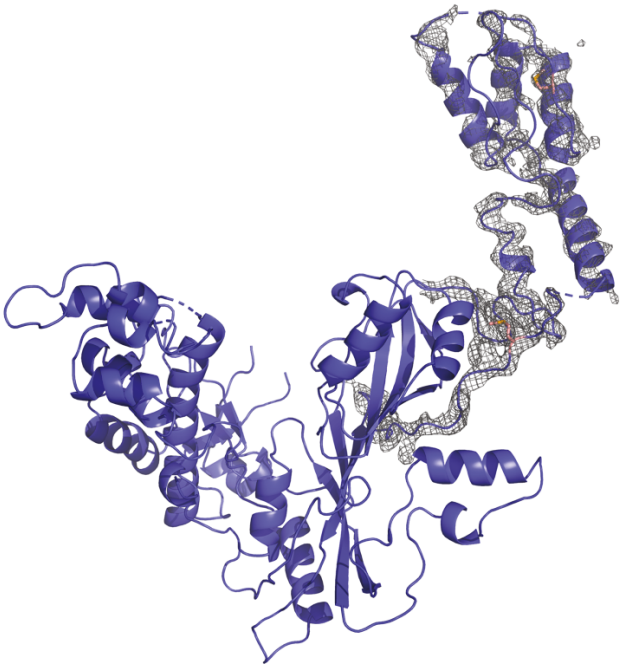


Figure S2: Comparison of Pus7 structures. (A) Rendering of the electrostatic surface potential of yeast Pus7 generated with ABPS Electrostatics (20). Negatively charged regions are shown in red, and positively charged regions are shown in blue. **(B)** Catalytic residue D256, Pus7 numbering, is shifted ~4A relative to the same residue in TruD (D80). Figure shows alignment of the yeast Pus7 active site (purple) and residues (gray) with the equivalent residues in TruD (yellow, PDB: 1SB7)(21). Pus7 numbering in black, TruD numbering in yellow. Distances measured both from CAlpha position and from the carboxyl on D256/D80. **(C)** Superposition of yeast Pus7 (light gray, blue) and human Pus7 (dark gray, yellow, PDB: 5KKP)(22), (C_{alpha} RMSD = 3.743 for 144 atoms) and rotated 180 degrees to show the difference in position of the insertions (I, II, and III) in yeast (blue) and human (yellow) Pus7. The catalytic residue D256, yeast numbering, is shown in the active site (light gray spheres). **(D)** Top down view of yeast and human superposition, looking down into the active site. **(E)** Putty representation of Pus7 colored according to B factors. Residues with the lowest B factors in blue (min = 20Å) and maximum in red (max=200Å). (**F)** 2Fo-Fc maps showing experimental electron density (gray mesh) around yeast Pus7 ID-I (blue) contoured at 1.5σ. Methionine residues (M88 and M145, orange), shown as sticks, used for SAD phasing.(**G**) Superposition (using 136 – 336 c-alpha atoms of the TRUD and PUS domains, RMSD: ~2.52Å for these domains) of TruD homologs, including: each molecule in the asymmetric unit of each *E.coli* TruD structure (gray, PDB: 1sb7, 1si7, 1szw), both TruD molecules in the asymmetric unit of the *Methanosarcina Mazei* structure (gray, PDB: 1z2z), the single Pus7 molecule in the human structure (yellow, PDB: 5kkp), and the single molecule in the yeast Pus7 structure reported here (blue).


A.
180°
-
-5.000
5.000
+
B.
D80
4.3
E31
D256
4.9
K21
K61
E71
N129
F27
N305
F67
F131
F307
C.
I
II
180°
I
II
III
D.
II
I
III
E.
I
II
III

F



G

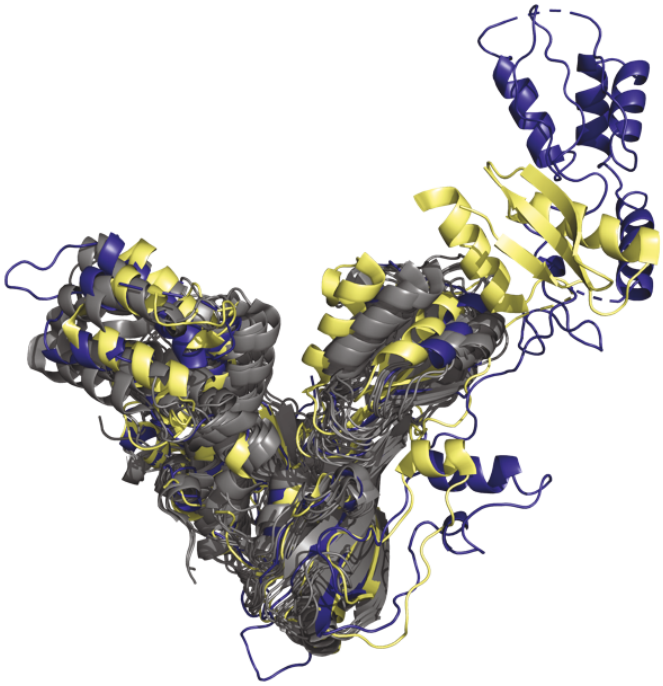


Figure S3: Sequence alignment of representative TruD family members.

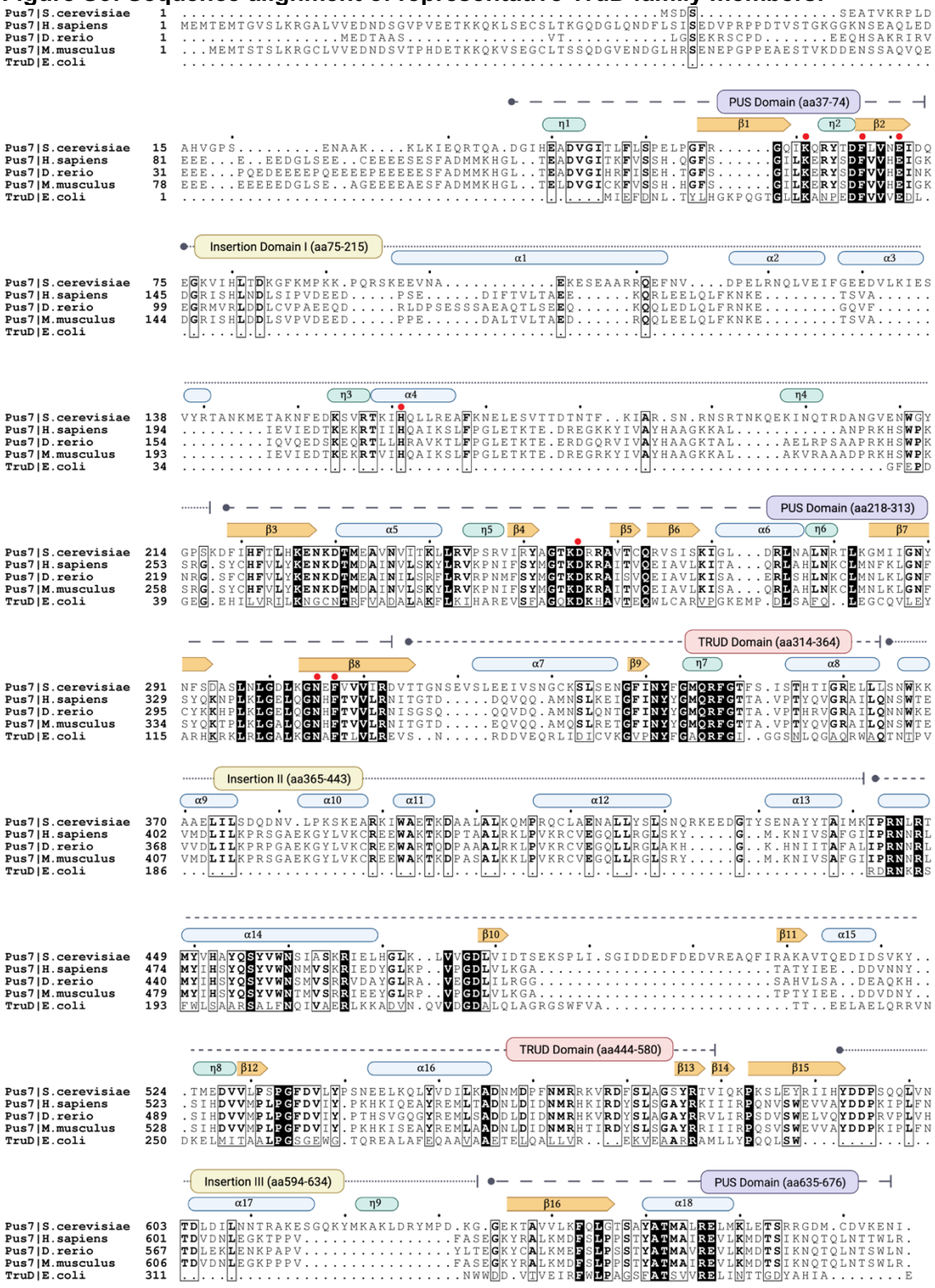


Figure S4: ID-I contains a single strand nucleic acid binding R3H domain. DALI alignment of the R3H motifs (based on 56 atoms, RMSD: 2.788 Å) from humanPus7 ID-I (yellow, PDB: 5kkp) and PARN (blue, PDB: 2a1s).

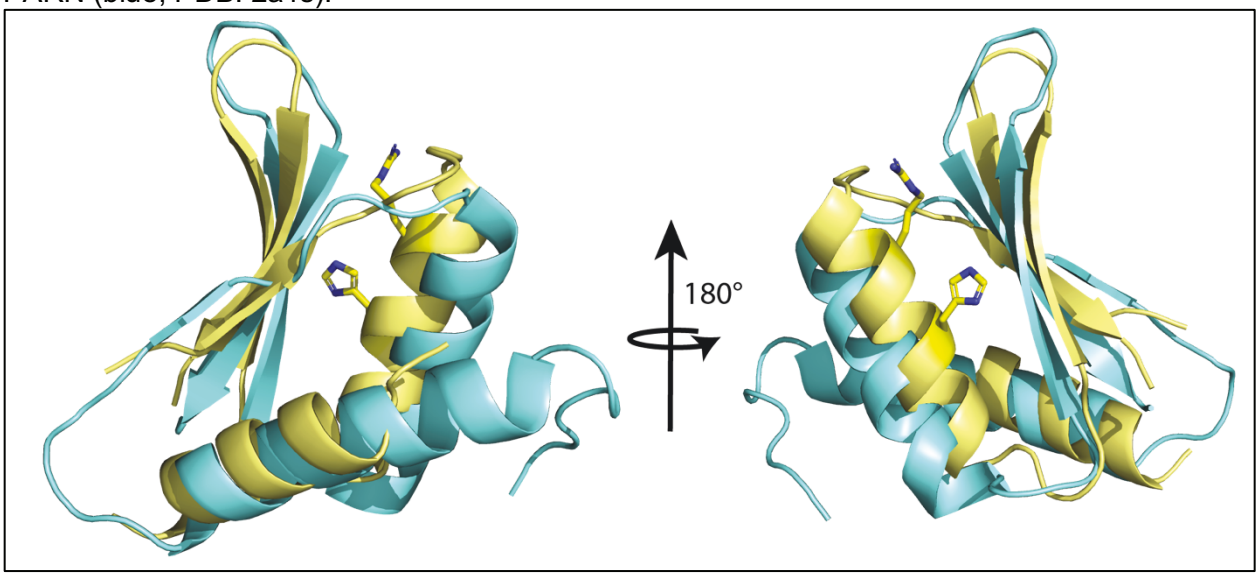


Figure S5: *S. cerevisiae* cell growth under different conditions.

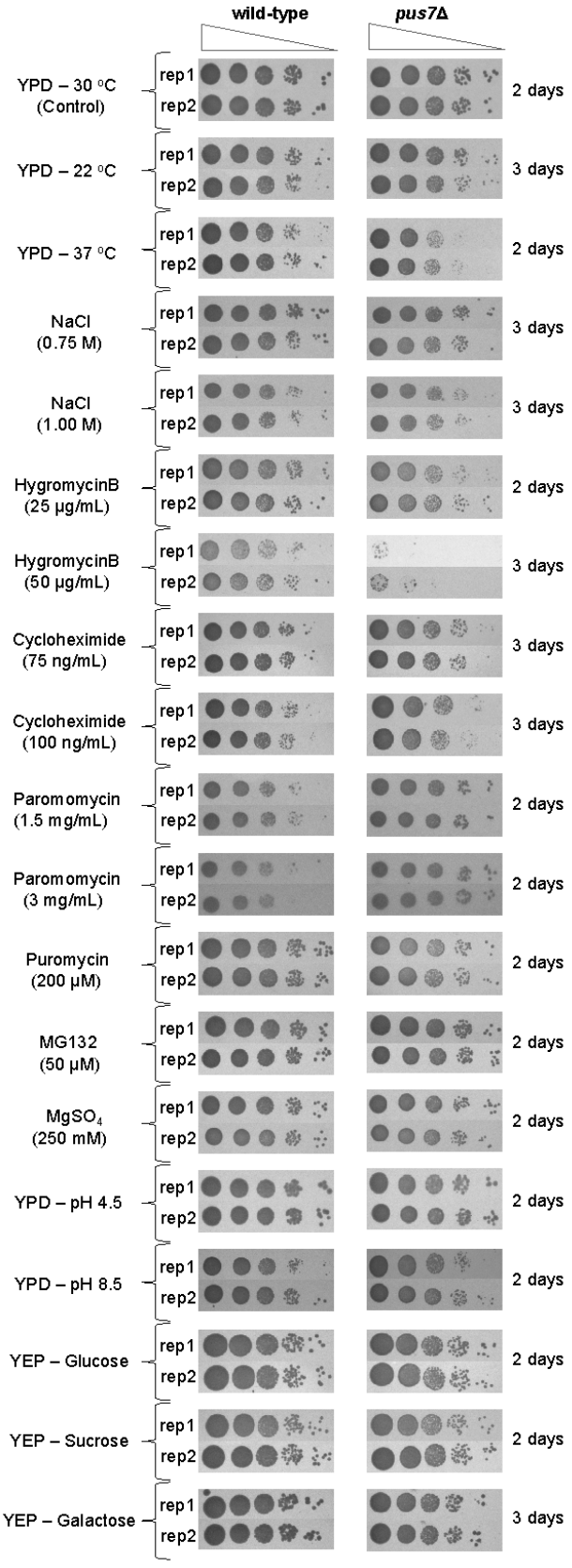


Figure S6: Example S. cerevisiae cell growth in liquid media. Growth curves for wild-type and *pus7*Δ cells in YPD at 30°C after the addition of **(A)** nothing, **(B)** cycloheximide, and **(C)** paromomycin.

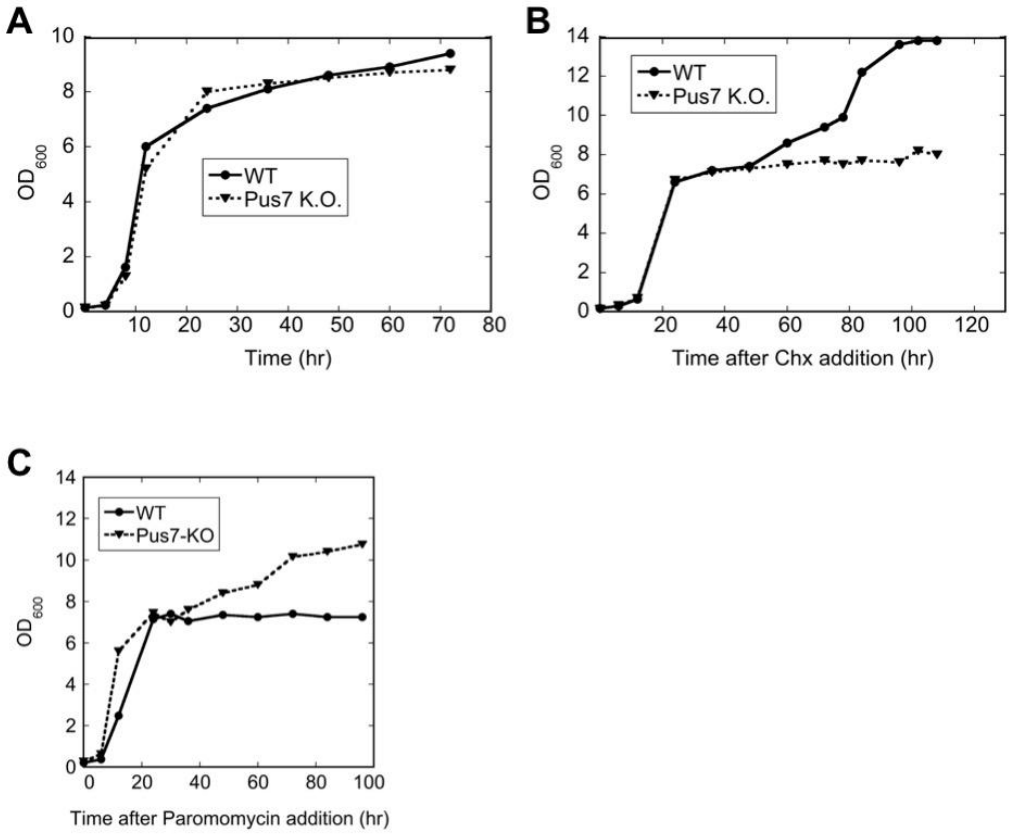


Figure S7: Ribosome occupancies are affected in the absence of Pus7. (A) Ribosome protected footprint (RPF) read length distribution. Distribution of ribosome protected fragments (RPFs) length show that most of the RPFs are between 27-30 nucleotide length. **(B)** ~50-60 % of these RPFs are in-frame (frame 0) **(C)** Ribosome occupancies are altered in p*us7*Δ compared to wild-type cells. Fold change in the ribosome codon occupancies was simply calculated by dividing the number of mapped RPFs in the P-site of *pus7*Δ to wild-type.

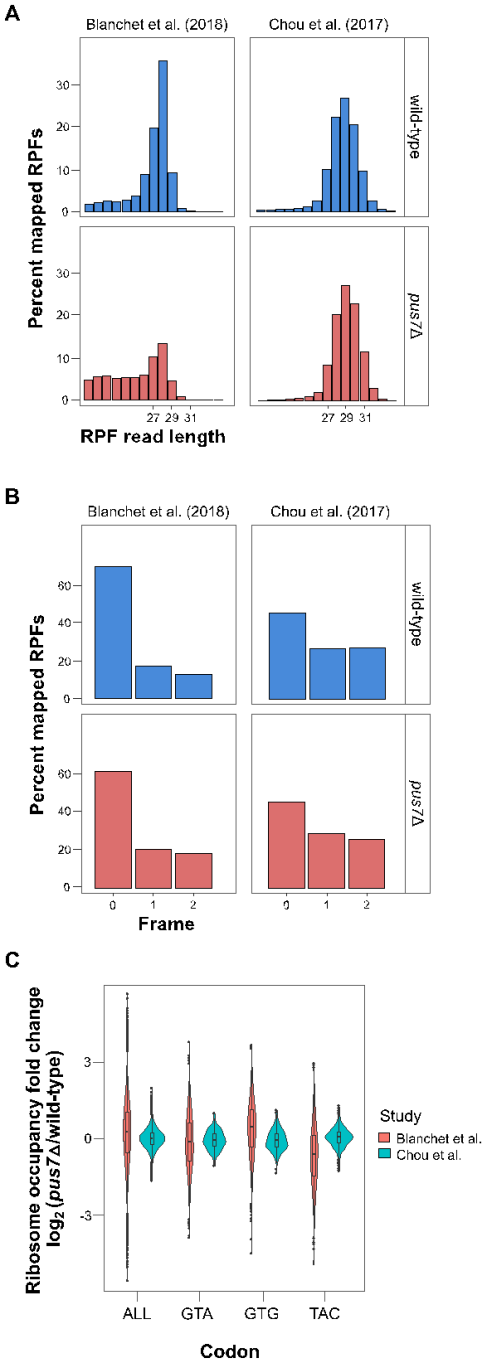
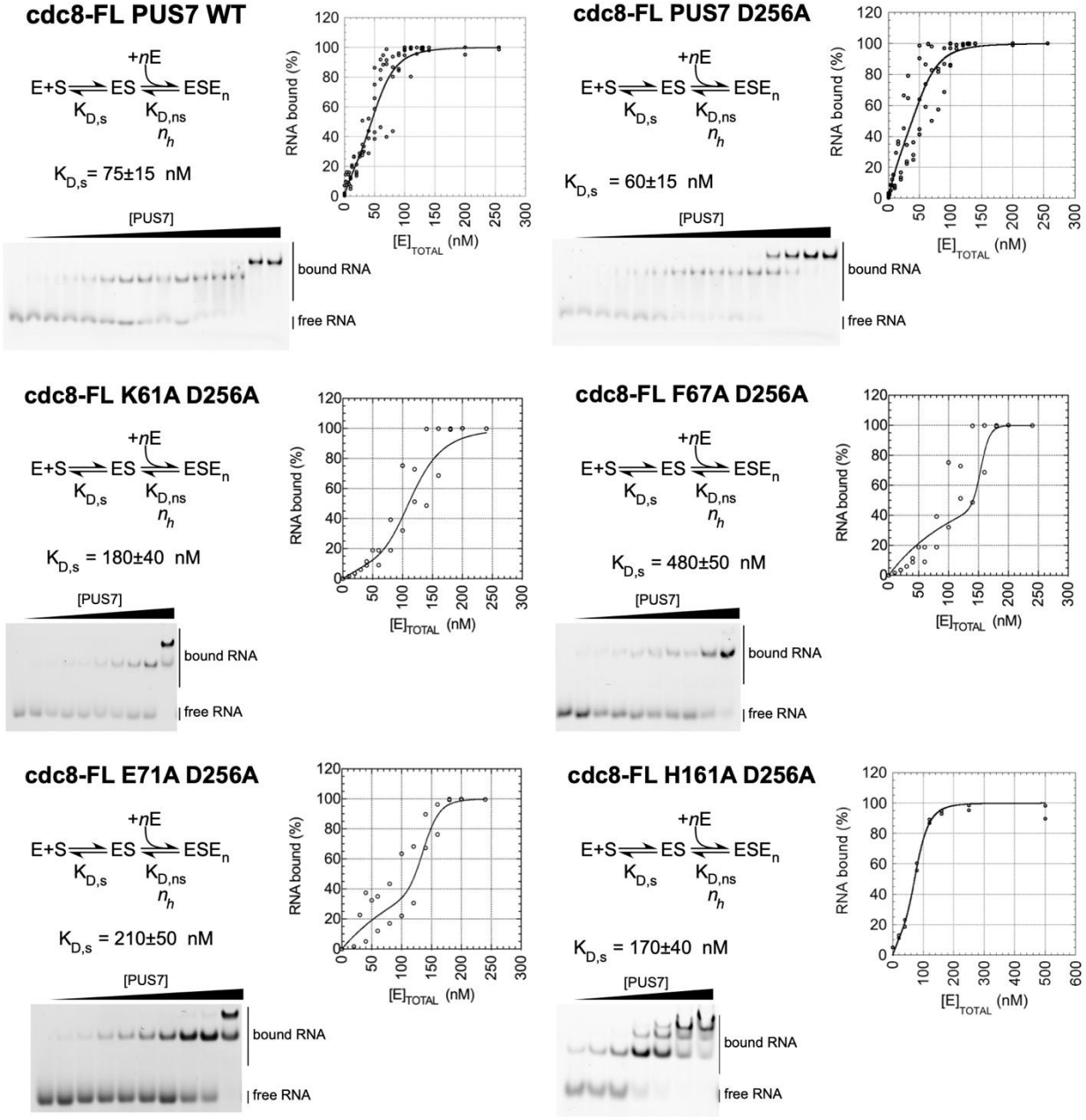
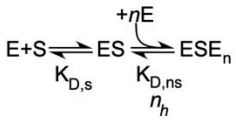


Figure S8: Raw EMSA data

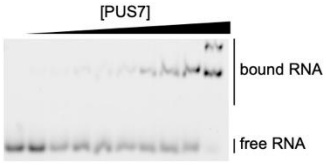
Each panel is titled “SUBSTRATE PROTEIN” in bold text. Panels are grouped by substrate and then by protein mutations. Each panel shows the binding model used for curve fitting, one gel image, and a single curve fitted to all replicate data sets. The dissociation constant for the specific binding step of the model is noted along with the error of the fitted parameter.

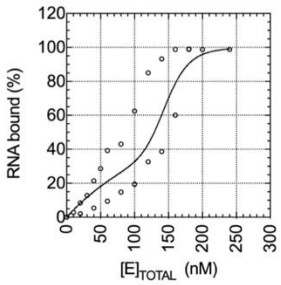


cdc8-FL N305A D256A

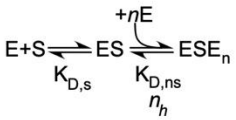


$K_{D,s}$ = 230±60 nM

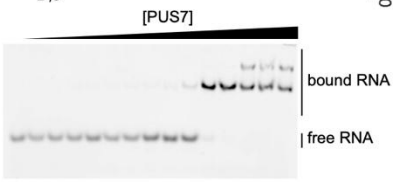


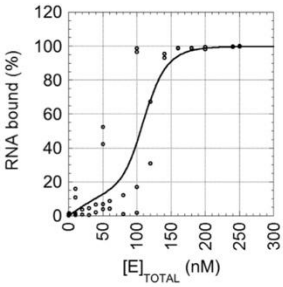


cdc8-FL F307A D256A

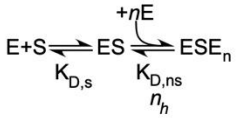


$K_{D,s}$ = 340±170 nM

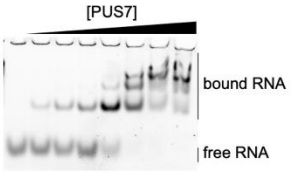


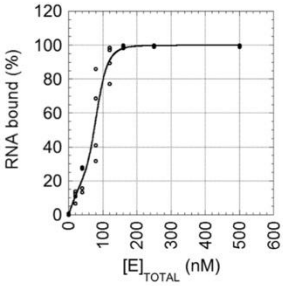


cdc8-FL ΔID1 D256A

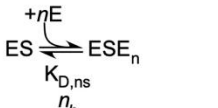


$K_{D,s}$ = 160±40 nM

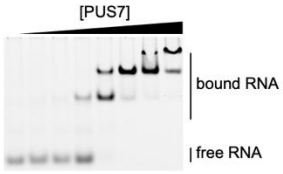


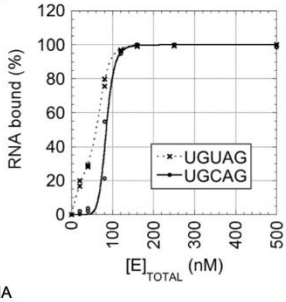


cdc8-FL nontarget D256A

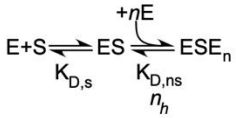


nontarget: UG**C**AG
target: UG**U**AG

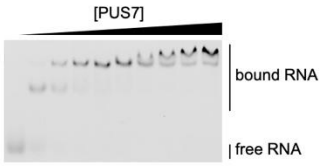


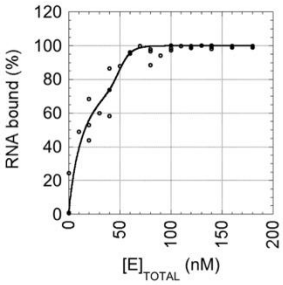


cdc8-A PUS7 WT

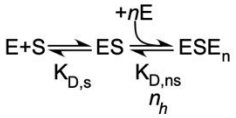


$K_{D,s}$ = 15±2 nM

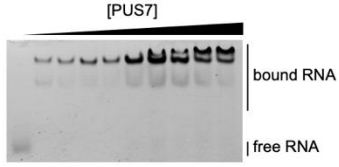


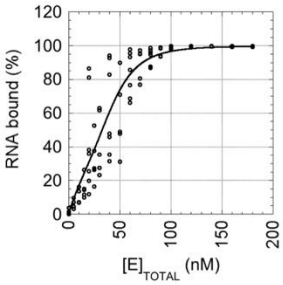


cdc8-A PUS7 D256A

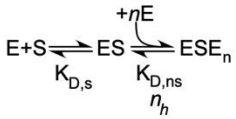


$K_{D,s}$ = 57±4 nM

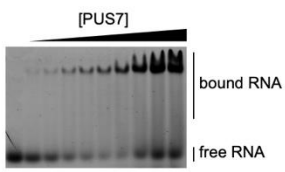


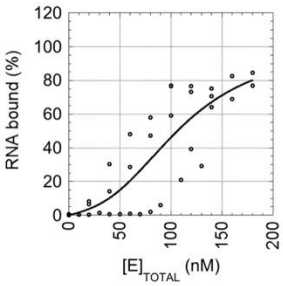


cdc8-B PUS7 WT

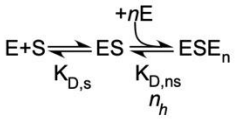


$K_{D,s}$ = 630±1500 nM

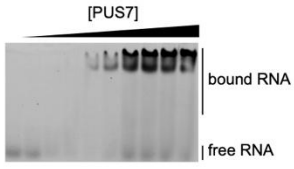


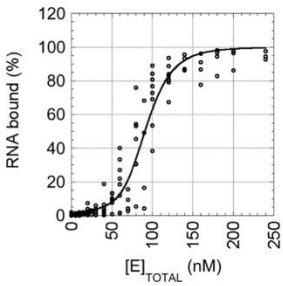


cdc8-B PUS7 D256A

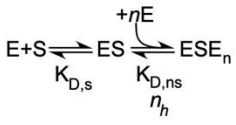


$K_{D,s}$ = 800±320 nM

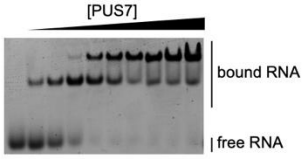


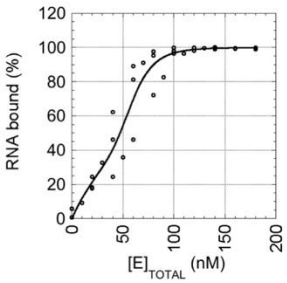


cdc8-C PUS7 WT

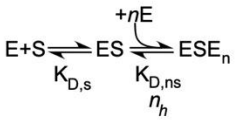


$K_{D,s}$ = 74±18 nM

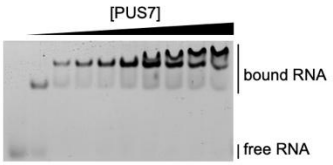


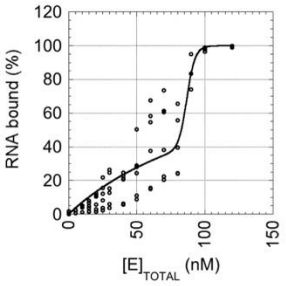


cdc8-C PUS7 D256A

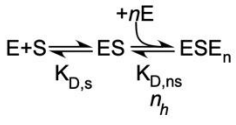


$K_{D,s}$ = 130±13 nM

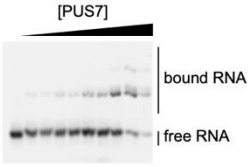


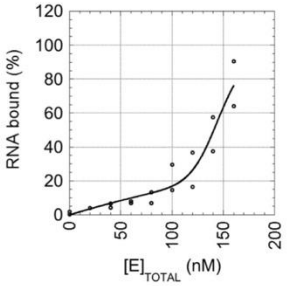


tArg PUS7-WT

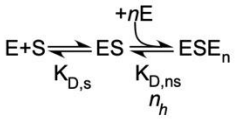


$K_{D,s}$= 550±160 nM

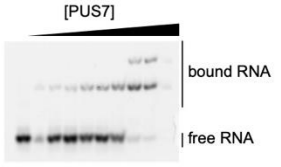


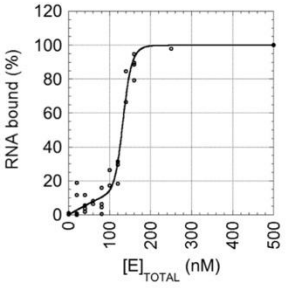


tArg PUS7-D256A

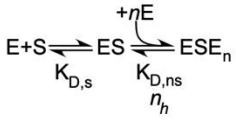


$K_{D,s}$= 700±170 nM

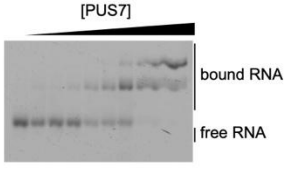


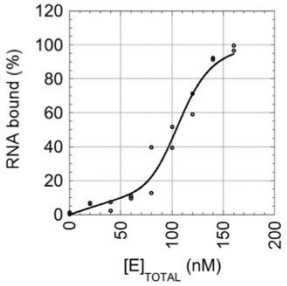


tArg-C13U PUS7-WT

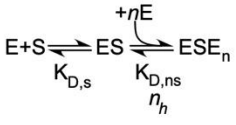


$K_{D,s}$= 470±180 nM

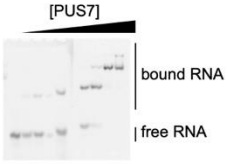


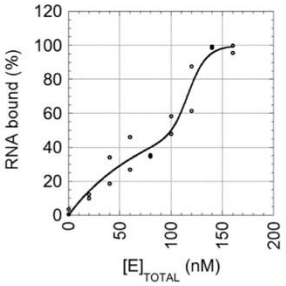


tArg-C13U PUS7-D256A

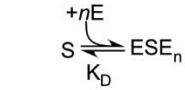


$K_{D,s}$ = 120±20 nM

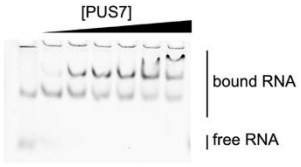


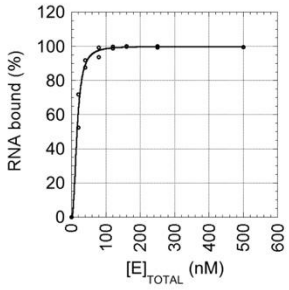


tAsp PUS7-D256A

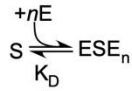


K_D = 16±1 nM

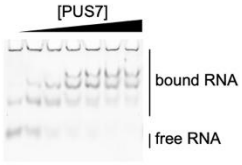


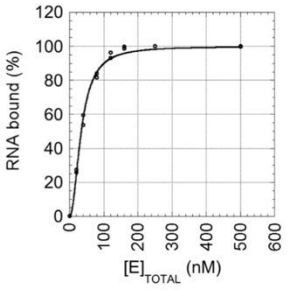


tAsp PUS7-ΔID1 D256A

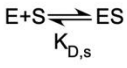


K_D = 34±1 nM

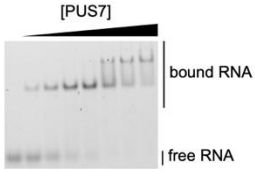


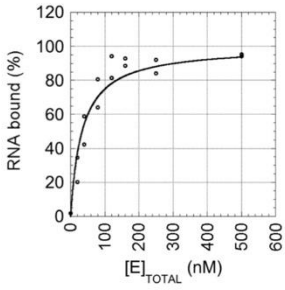


ST1 PUS7-D256A

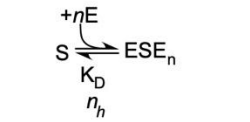


$K_{D,s}$ = 34±4 nM

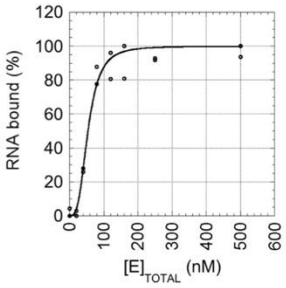


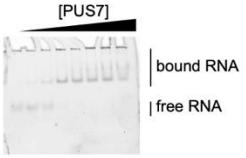


ST1 PUS7-ΔID1 D256A



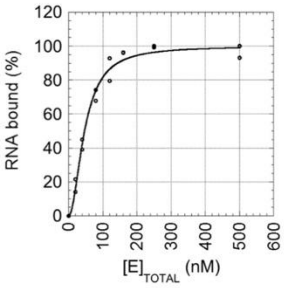
K_D = 54±3 nM

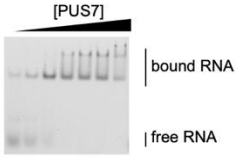




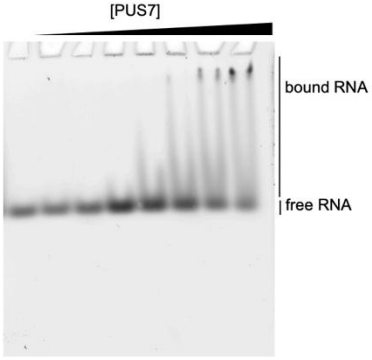
SNT1 PUS7 D256A

K_D = 47±2 nM



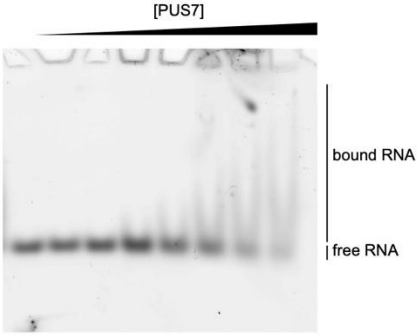


ST2 D256A



binding not quantitatively analyzed

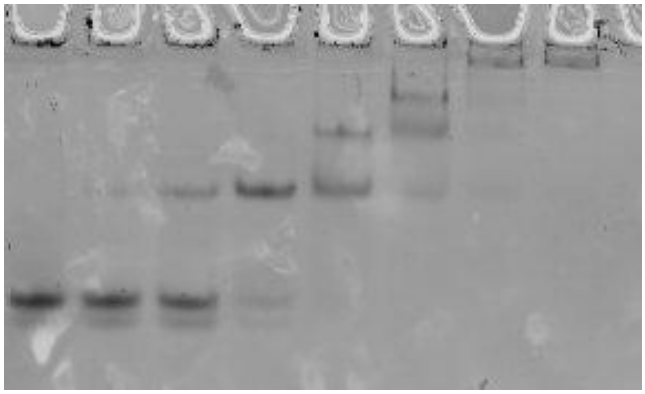
ST2 PUS7-ΔID1 D256A



binding not quantitatively analyzed; however there is less bound (smeared) RNA with ΔID1 relative to D256A alone

Figure S9: Yeast PUS1 nonspecifically binds RNA and catalyzes pseudouridinylation outside its consensus sequence. A. EMSA using PUS1 and its GLK1 target RNA showing specific and nonspecific binding events. B. Measurement of pseudouridine synthase activity on a variety of PUS7 and PUS1 substrate RNAs. The tRNA substrates are positive controls and show the expected pattern of activity. The MFKKX substrate contains two UGUAG motifs and mutation of one of them eliminates pseudouridinylation at that site by both PUS1 and PUS7.

A



B

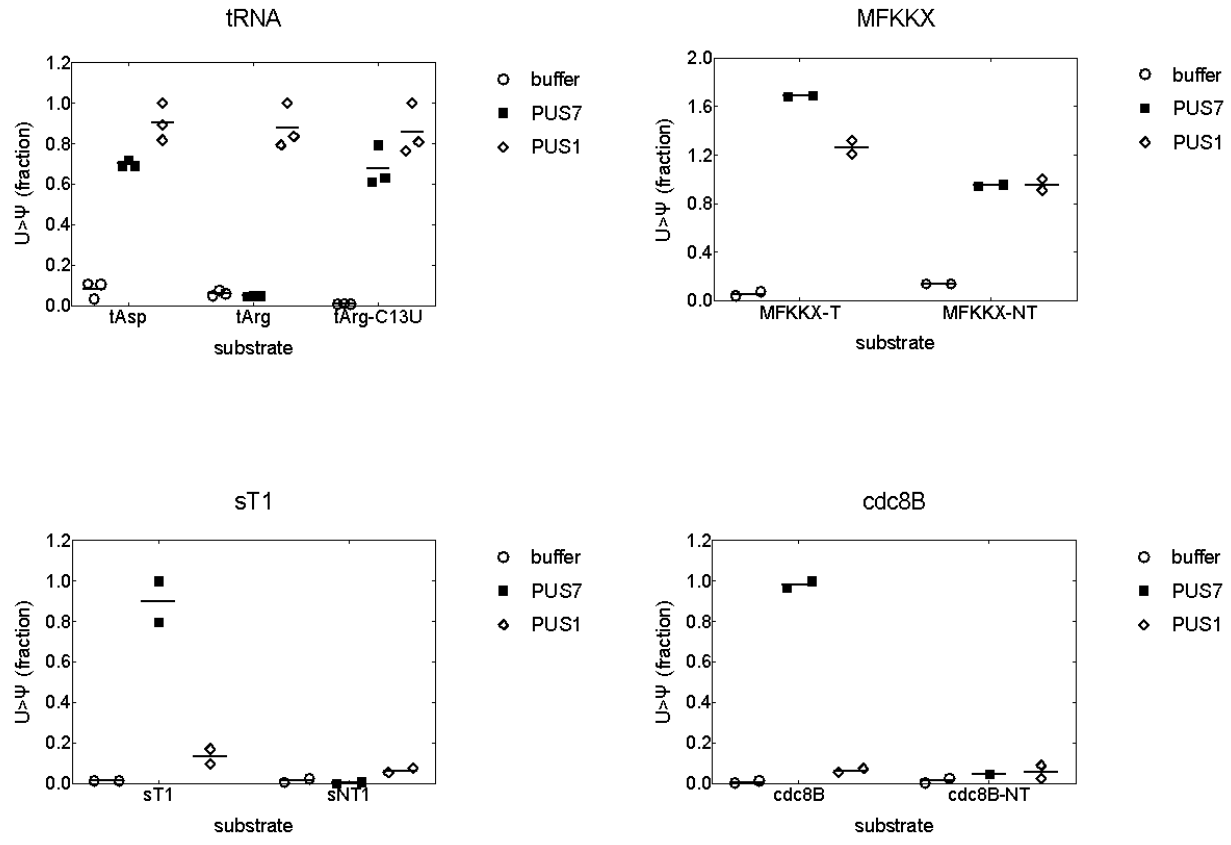


Figure S10: Models utilized in analysis of EMSA data. (A) Simultaneous binding of *n* PUS7 enzymes to a single RNA substrate, otherwise known as a Hill binding curve. This model was used when no specific binding was apparent. Binding curves were fit to the equation shown. The EMSA assay allowed direct estimation of free enzyme concentrations, so we fit the data using both free and total enzyme concentrations. The differences between these fits was much smaller than the difference in fits of data from independent replicates. Since using total enzyme as the independent variable was not the limiting factor in the precision of our measurements, we used total enzyme as the independent variable for simplicity. (B) Model for binding of one PUS7 enzyme to a single specific site on the substrate RNA, followed by simultaneous binding of *n* PUS7 enzymes to *n* non-specific sites on the same RNA. This model was used to fit data when the Hill equation underestimated the fraction bound at lower concentrations of enzyme, reflecting the existence of a unique site with a lower K_D for PUS7. (C) A realistic model for binding of one or more PUS7 enzymes, in arbitrary order, to a single specific site and one or more nonspecific sites on a single RNA. Occupancy of nonspecific sites is indicated by superscripts *i*, *j*, *k*, … on the S. Nonspecific sites can be bound in any order (e.g., *k,l,l,j*) but are depicted in alphabetical order for convenience.

A

$$S_f + n\,E_f \overset{K_{D,app}}{\rightleftharpoons} S E_{NS,n} \qquad \theta_b \approx \frac{E_T^{n_H}}{K_{D,app}^{n_H} + E_T^{n_H}}$$

B

$$S_f + E_f \overset{K_{D,app1}}{\rightleftharpoons} E_S S + n\,E_f \overset{K_{D,app2}}{\rightleftharpoons} E_S S E_{NS,n} \qquad \theta_b \approx \frac{E_T \bullet \left(K_{D,app2}^{n_H} + E_T^{n_H}\right)}{K_{D,app1} + E_T \bullet \left(K_{D,app2}^{n_H} + E_T^{n_H}\right)}$$

C

$$\begin{array}{ccccccccc} S_f & \rightleftharpoons & S^{i}E_{NS,1} & \rightleftharpoons & S^{i,j}E_{NS,2} & \rightleftharpoons & \cdots & \rightleftharpoons & S^{i,j,k,\ldots}E_{NS,n} \\ \updownarrow & +E_f & \updownarrow & +E_f & \updownarrow & +E_f & & +E_f & \updownarrow \\ E_S S & \rightleftharpoons & E_S S^{i}E_{NS,1} & \rightleftharpoons & E_S S^{i,j}E_{NS,2} & \rightleftharpoons & \cdots & \rightleftharpoons & E_S S^{i,j,k,\ldots}E_{NS,n} \end{array}$$

Figure S11: Stopped flow assessment of binding kinetics. (A) Experimental set-up, as described in the corresponding SI Appendix Methods. **(B)** Stopped-flow traces of Fl-CDC8 rapidly mixed with 0, 20 and 750 nM of D256A Pus7 protein. **(C)** Traces at higher D256A concentrations were biphasic. This shows a 750 nM trace fit with one or two phases. **(D)** All of the $k_{obs,1}$ values measured are plotted as a function of D256A Pus7 concentration.

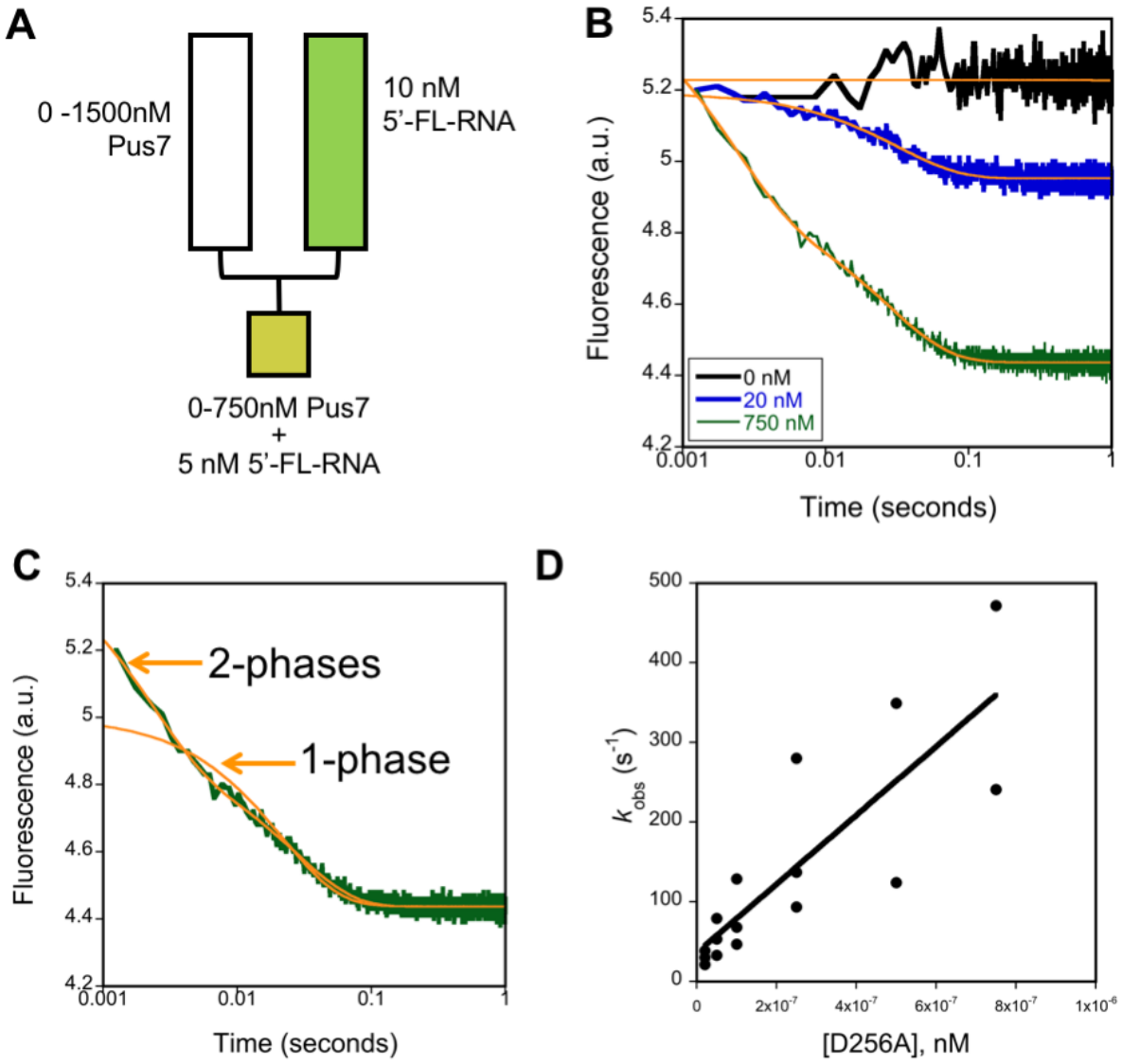


Figure S12: Deletion of ID1 does not broadly affect pseudouridinylation of total RNA in vitro. Total cellular RNA extracted from Δpus7::kanMX was pseudouridinylated in vitro using PUS7FL or PUS7ΔID1. Pseudouridinylation of known sites was assayed using CLAP (25).

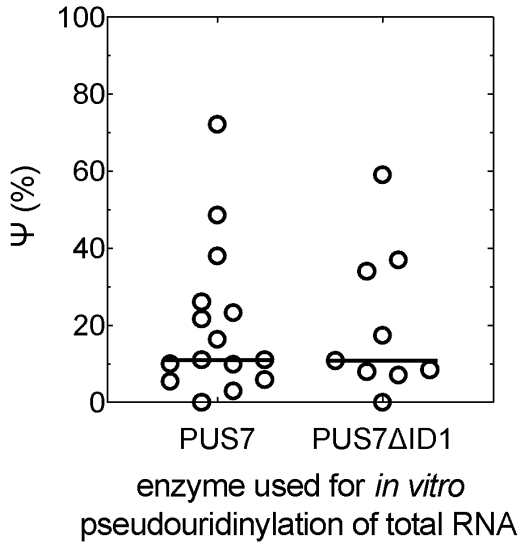


Figure S13: Deletion of ID-I influences pseudouridylation efficiency in a target dependent manner.

Total cellular RNA was extracted from *pus7::kanMX* yeast and pseudouridinylated in vitro with PUS7fl or PUS7ΔID1 (left side) or extracted from *pus7::kanMX* yeast expressing PUS7FL or PUS7ΔID1 (right side). Pseudouridinylation was assayed at specific sites using CLAP (25). The difference between mean pseudouridinylation level at sites in RNA exposed to PUS7ΔID1 and RNA exposed to PUS7fl is shown on the y axis.

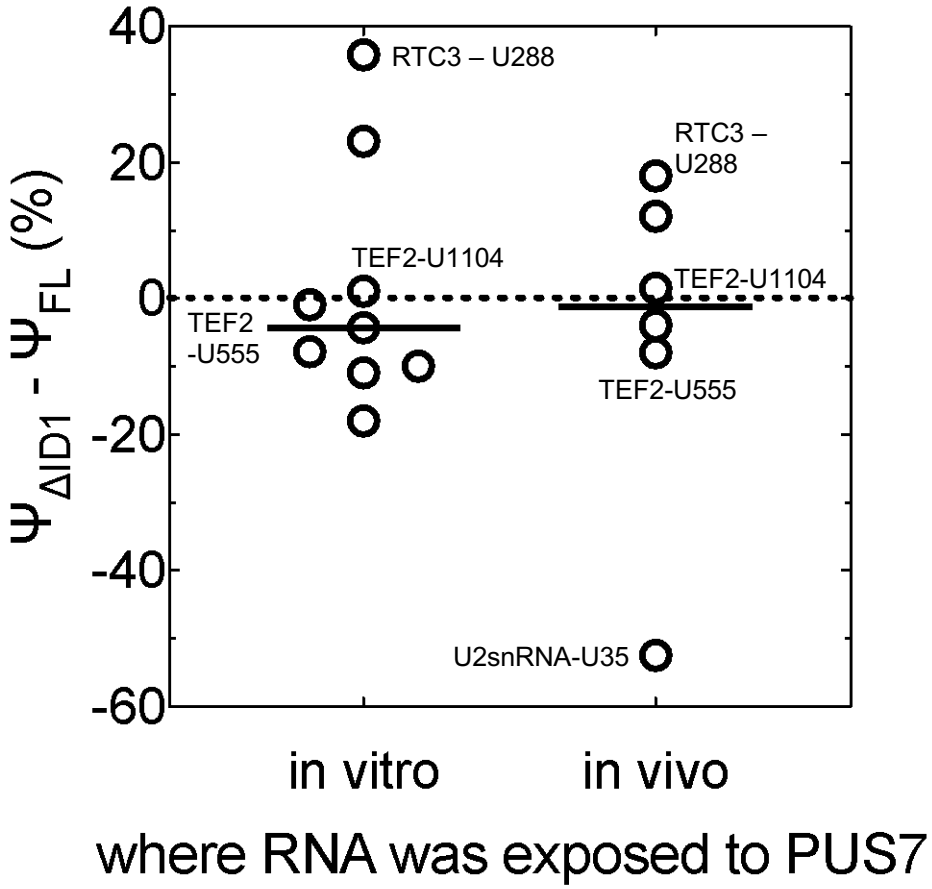


Figure S14: Expression of PUS7ΔID1 confers no obvious phenotypic defects relative to PUS7FL. PUS7 was expressed from a CEN plasmid under the control of a GPD promoter in WT and *Δpus7::kanMX* yeast and assayed by spot plating under the indicated conditions. Three independent transformants were assayed for each plasmid.

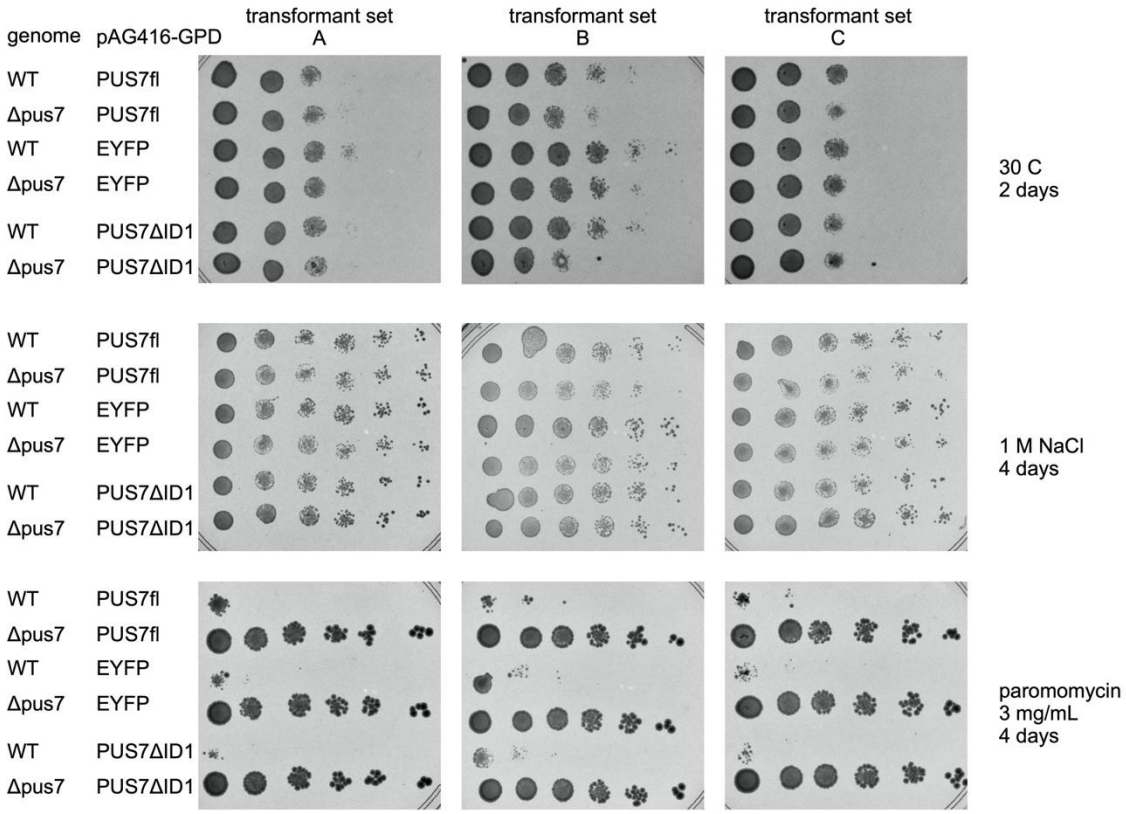


Figure S15: Isolation of PUS7fl and PUS7ΔID1 expressing clones. Three independent transformants were isolated for each strain/plasmid combination.

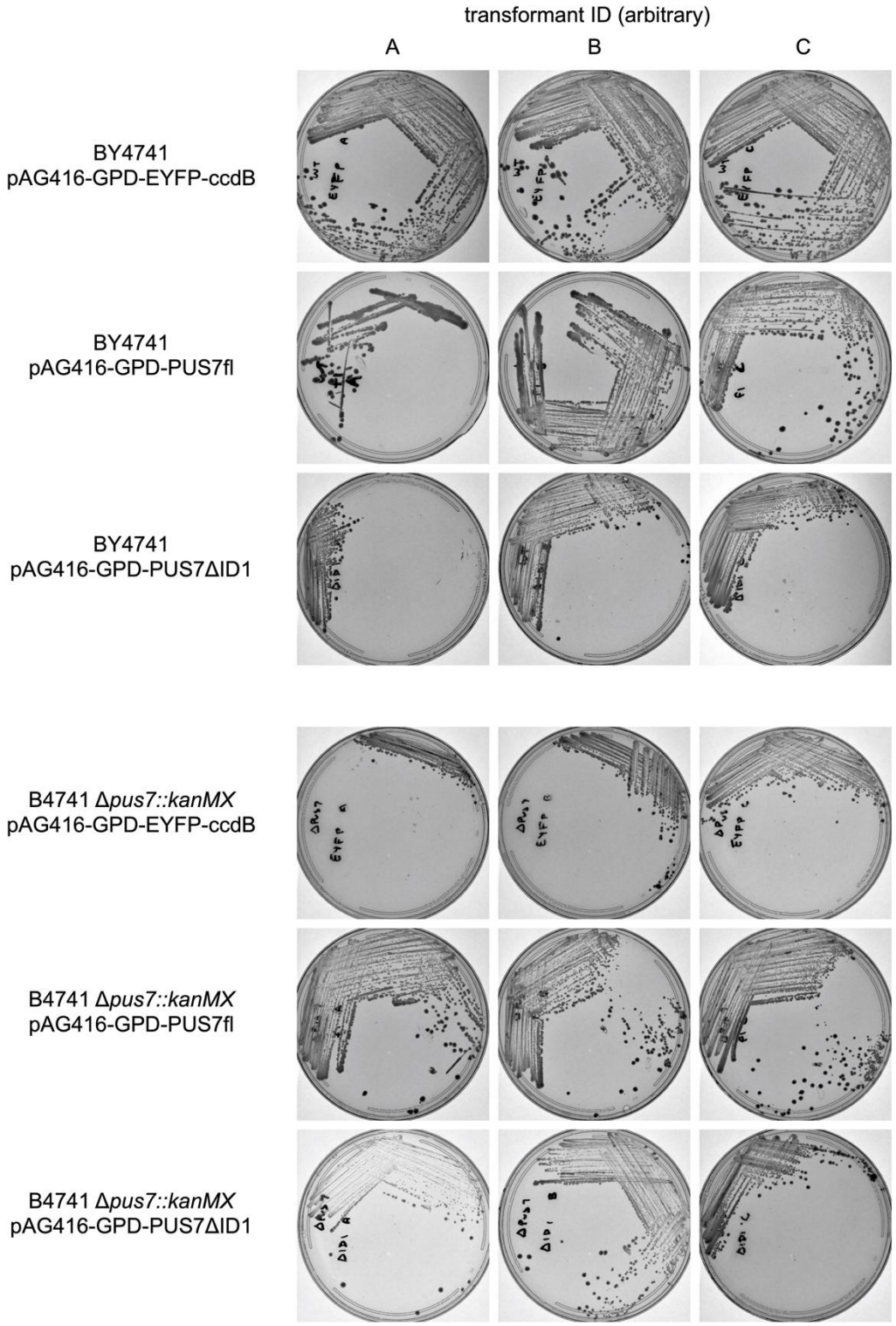


Figure S16: Secondary structure prediction of Pus7 modified sites in mRNA coding regions reported in Carlile, et al. Nature (2014) (23).

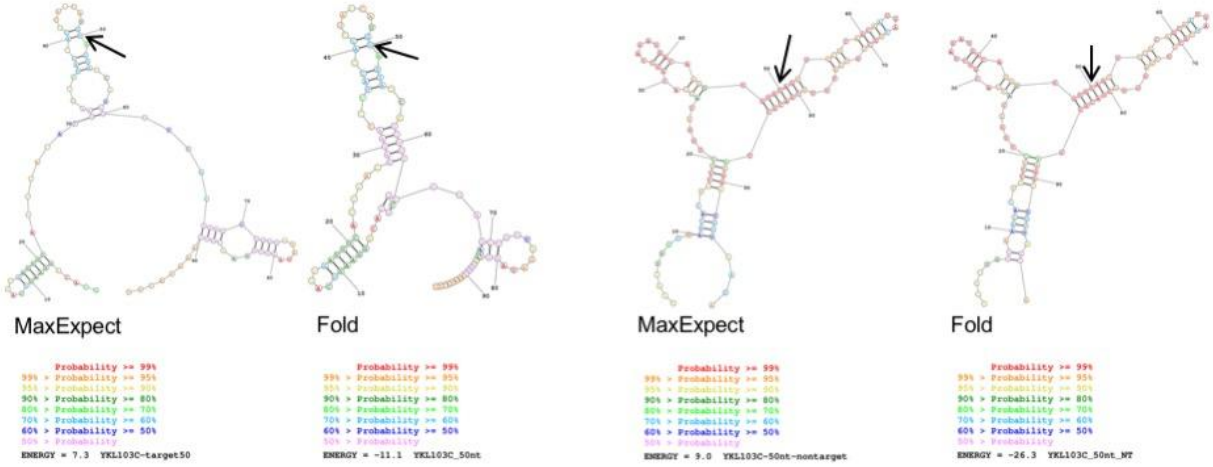
YKL103C

Target UGUAG

GCACAGGAAUUCAUUGAUUUCAUUUACAAGAACC
CUACCACUUACCA**UGUAG**UAUCAUUUUUCGCGG
AGCUGUUAGAUAAGCAUAACUUCAAAUACUU

Non-target UGUAG

GUAUUGGUAAACACUUCCUUUUUGCACCACGUCU
AGAUGACAGGUUG**UGUAG**UUUCGCAGCAAUGAUU
GCUUUGAUUUGCUACGCUAAGGAUGUUAAUA



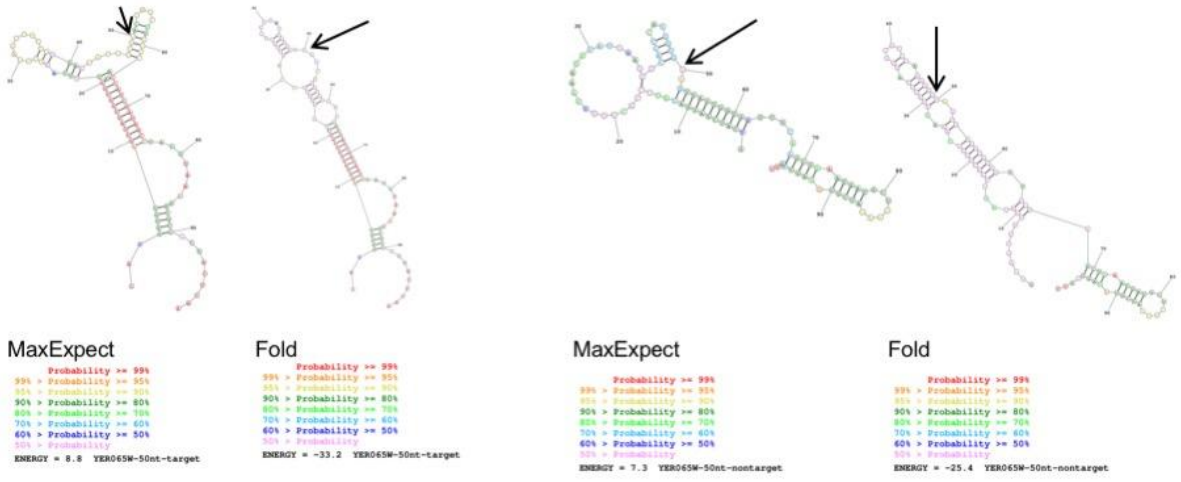
YER069W

Target UGUAG

CAUGGUGUAGUUAUUGGUGGUUUCAAGCUGAAU
GAUGCUGAAGAUCG**UGUAG**UUGUUUGCGCAACC
AUCGAUAACUUACUUAAAGGCGCCGCUACUCAA

Non-target UGUAG

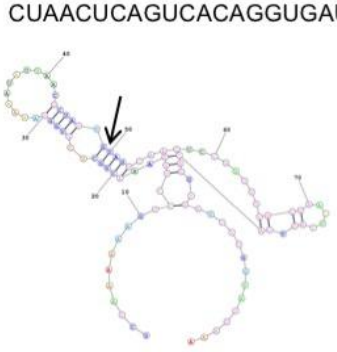
GUCAUCGAUGAUAUCCCACUGGUUAAAGAUAUUG
AGGGCACCCAUGG**UGUAG**UUAUUGGUGGUUUCAA
GCUGAAUGAUGCUGAAGAUCGUGUAGUUGUU



YMR062C

Target UGUAG

UCGAAAAAAGUCCUUGAAACUGCUCGUGGUAAAA
AUAUUAACGCUAU**UGUAG**UCAAUUCCGGUUGUG
CUAACUCAGUCACAGGUGAUCUUUGGUAUGAAA



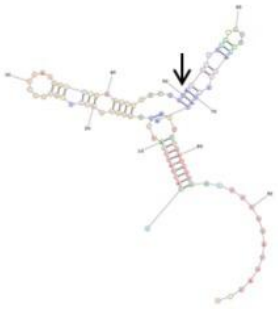
MaxExpect

Fold

YPR176C

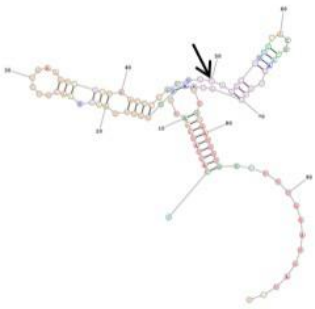
Target UGUAG

GCACUAUCAAUUUUGGGUGAAUUAACGUCUGAA
GUUGUUGACCCUGC**UGUAG**ACUUUGUACUCAAG
UGUUAUAAUUUUGAUGGUGGCUUUGGAUUAUGU



MaxExpect

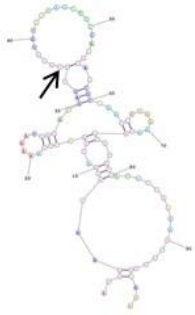
Fold



YNL336W

Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA



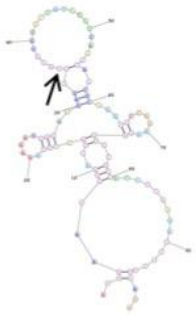
MaxExpect

Fold

YBR032C

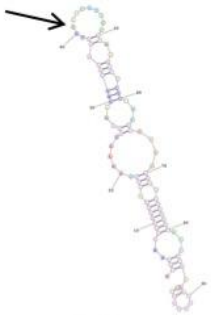
Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA



MaxExpect

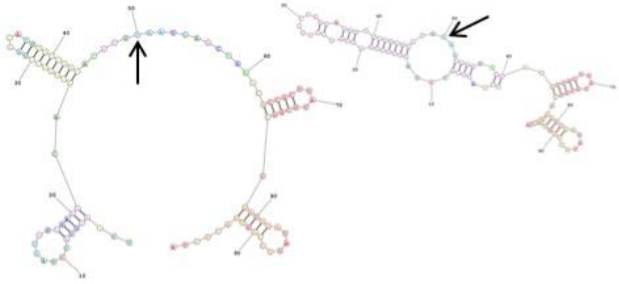
Fold



YDR380W

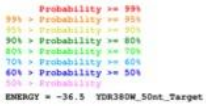
Target UGUAG

UCUAUUCACCUAGUGUUGAAUCAGCUGGCCUAA
GAUGGGUCGGCACG**UGUAA**UGAACUGAACGCCG
CUUAUGCGGCCGACGGAUAUUCCCGUUACUCUA



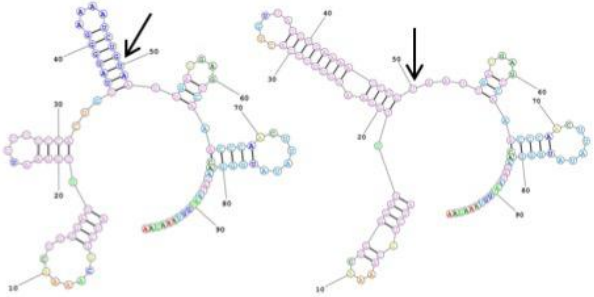
MaxExpect

Fold



Non-Target UGUAA-1

AUCUGCAAAACUGGGAUUUGGAAUUUUUCCACU
GUUAUGGGAAAAUC**UGUAA**UUGAUGAGUCAAAC
CCAACUUAUAUGGGUCAAUAUAAUGGUAAAGAA

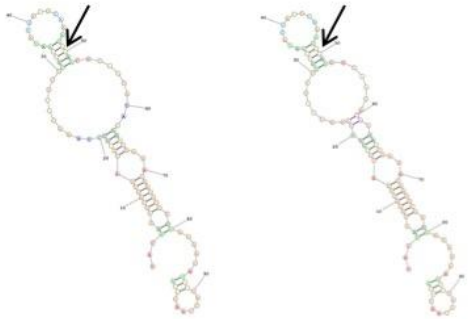


MaxExpect

Fold

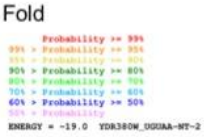
Non-Target UGUAA-2

UACAAGCGCAUUGACGUUUCUAAACUUUCUUUG
CAAUAUGAUUCAAA**UGUAA**CUCAAUAUACGAACG
AAACAAUGCGGUUAGAAGAUCCUACCAAUGGA



MaxExpect

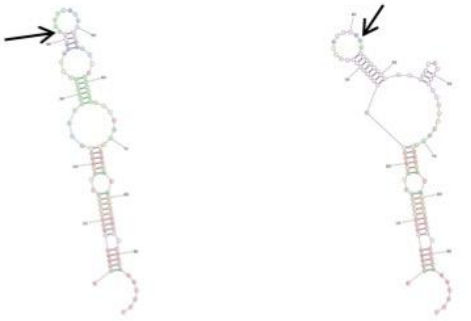
Fold



YNR028W

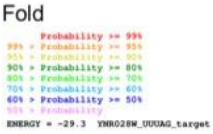
Target UUUAG

AUGAAGAGCUUUUUUCUUUAUCUUUAUGUGGCA
UUCAUG**UUUAG**UUGCAUAACGGCUCUGCCAUUA
CCUGUGGAUAACAAAAGGGCUUCUUCAGACUCC



MaxExpect

Fold



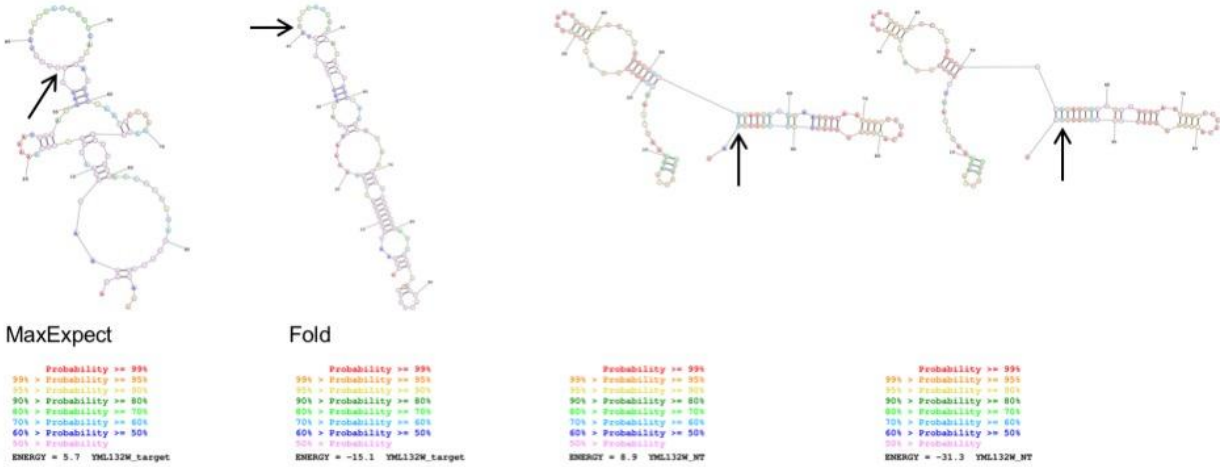
YML132W

Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA

Non-Target UGUAG

UCUGCGAAGAAAUCUAUGUGGCCUUUACCAUUG
AAUGUGGAACUAUGGCCAUACAUUAAAGAAGCG
CAAUUAUCCCGCAGUGAGGUGCUCUUAG**UGUAG**



YJR161C

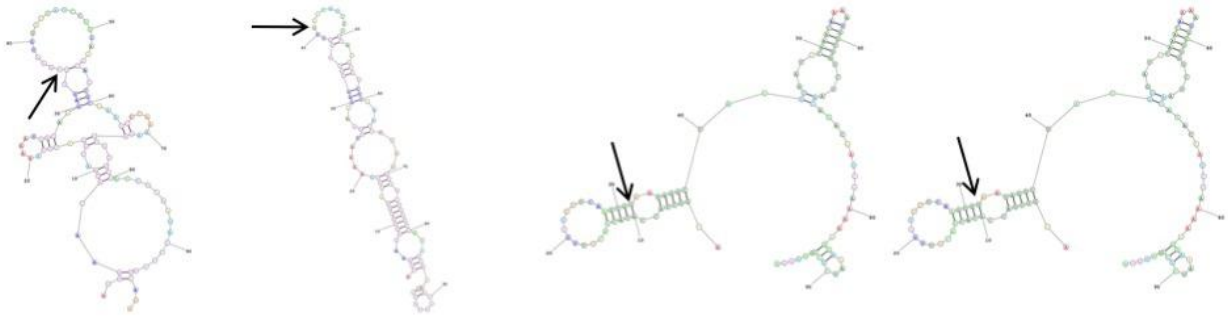
Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA

YBR036C

Target UGUAA

AUGUCUACCACACUACUUUGGUUUUCAAG**UGUA**
AUAGGCUACGUGAUUCAAACAAAAUGUUUGUCU
AACAUACAAUCUAAAAAGGAAAUCUCCGUGGGG



YPR124W

Target AUUAU

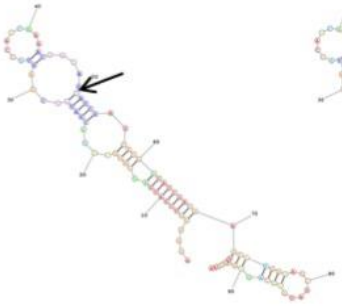
CAGUAGCUCUUCAGGGAUGGAUAUGGACAUGAG
UAUGGGAAUGAACU**AUUAU**CUGACUCCCACAUA
UAAAAACUAUCCAGUUUUGUUUCACCAUUUGCA

MaxExpect

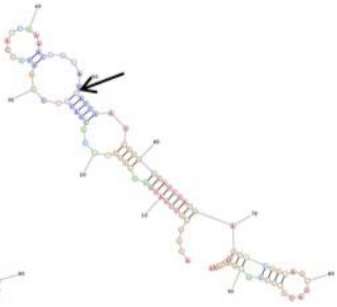
Fold

Non-Target AUUAU - 1

AUGGCGUCGAUGAGCAUGGAUGCGAUGUCUAGU
GCCAGCAAAACGAU**AUUAU**CGAGCAUGUCAUCG
AUGAGCAUGGAAGCGAUGUCCAGUGCGAGCAAA



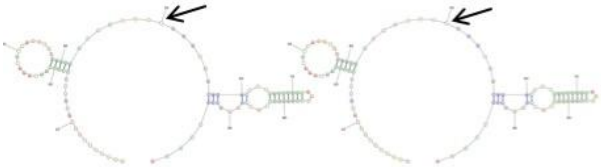
MaxExpect



Fold

Non-Target AUUAU - 2

UAAAUAUUCCACUUUACCUUCAGCAAACUCCAAA
GACGAAGGAAAAC**AUUAU**GACACAGAGAAUAAUU
UUGAAAUUCAAGGUUUACCUAAGCUGCCAAA

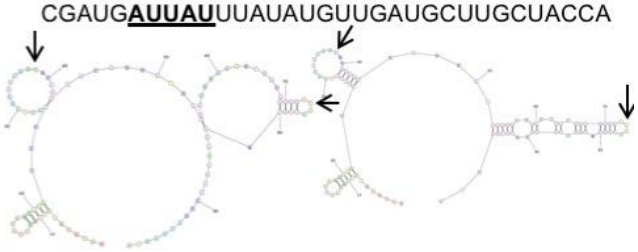


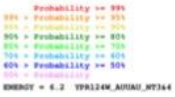
MaxExpect

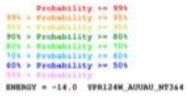
Fold

Non-Target AUUAU – 3 & 4

UAUUUGUUCCAUCUUUAAUGGAUCUCUUUCAUG
AC**AUUAU**AAGGGCGUUCUUAGUAUUUACCUCUA
CGAUG**AUUAU**UUAUAUGUUGAUGCUUGCUACCA



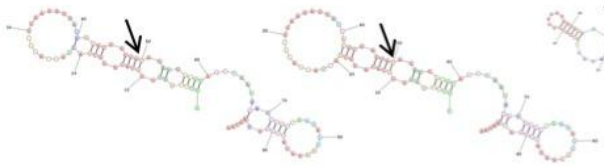




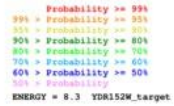
YDR152W

Target UGUAG

GGCGUGGUGGGCACAGGCGACGAGUAUAUGGAA
GAGGAUGACGCGAG**UGUAG**AUGACGUAGCCAAG
GGACUUGCCAAGACCGAAAUAGCAAAUCAAUAA



MaxExpect

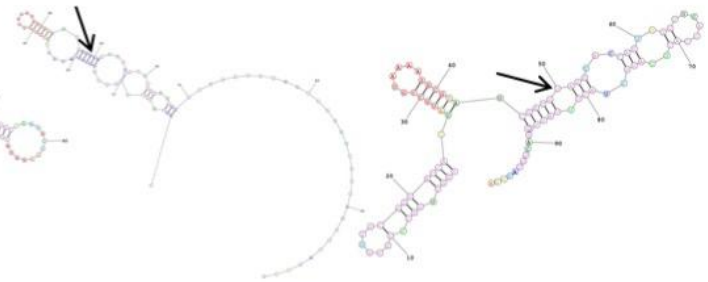


Fold

YLR258W

Target UGUAG

UGGUUGGCCGGUGUUGCGUUACCAUUAUGCCG
UAAAAGGCGUAUCGA**UGUAG**UUACCAUUtgtUUCA
CCACUCAUGCUACUUUAUUGGGACGGUAUUUA



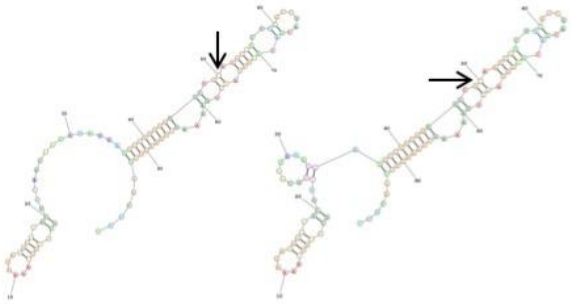
MaxExpect

Fold

YCL026B-C

Target UGUAA

GCCGGUCAAAAGAGGCCUGCUUCAGCAAGGGAU
GAGGCCUUUGGUUC**UGUAA**UCUUCUUCACCGAC
GACAAGGUAACUGAAAAGCUAAAGGCUGACUUC



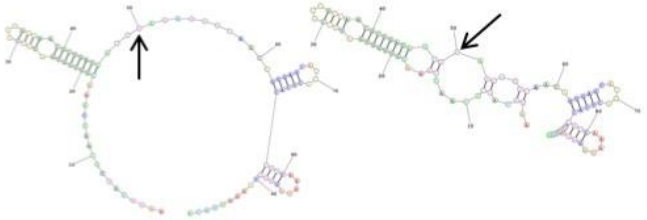
MaxExpect

Fold

YDR497C

Target UGUAG

AAGACAUCGCAAUCAAAUGUUGGUGAUGCCGUU
GGCAACGCUGAUAG**UGUAG**AGUUCAACAGUGAG
CAUGACUCACCUUCAAAGAGAGGUAAAAUUACA



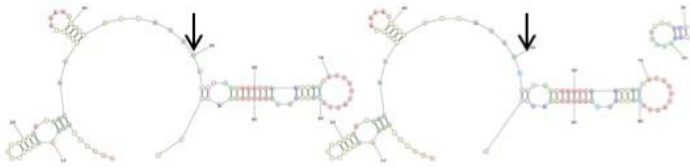
MaxExpect

Fold

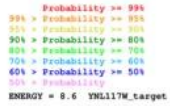
YNL117W

Target UGUAG

AUUAUCGUGAGACCCCGUGGUUGGCACAUGGUG
GAAAAGCACCUUUA**UGUAG**AUGAUGAACCAAUCA
GCGCUUCCAUCUUUGAUUUUGGUUUAUAUUUC



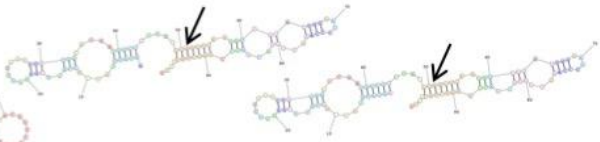
MaxExpect



Fold

Non-Target UGUAG

AAGAUGGAGCACCACUUGGAAGCUAAACUAUGG
AACGACGUCUUCUG**UGUAG**CUCAAGAUUACAUU
GGGAUCCCAAGGGGUACAAUCAGAGCUACUGUG



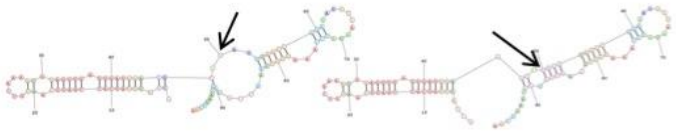
MaxExpect

Fold

YBR122C

Target UGUAG

GCUAUUUAUCACCAAUUCAAUGUCAAAAUGGAGU
UGAGUGAUGGAAG**UGUAG**UUAUCCGGAGAUCCC
AAUAUCCAAAGGGUGAAAUUAGAUUAAUUCAA



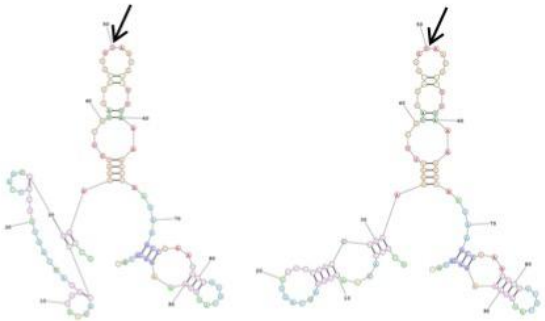
MaxExpect

Fold

YNR036C

Target UGUAA

GCUCCACAACUGGAUCAGUGUCCCCAGCGGAAG
GGUGUCGUAUUACG**UGUAA**UGGUUUUAAAGCCC
AAGAAGCCGAAUUCUGCGCAGAGAAAGGCGUGC



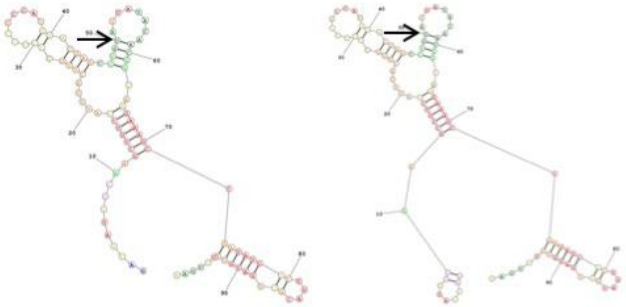
MaxExpect

Fold

YCL043C

Target UGUAG

GAUGAUUUCAAGCUUUCUAUUUACUUGCCCUCC
GCCAUGGACGAGCC**UGUAG**UAUACAACGGUAAG
AAAGCCGAUAUCGCUGACGCUGAUGUUUUUGAA



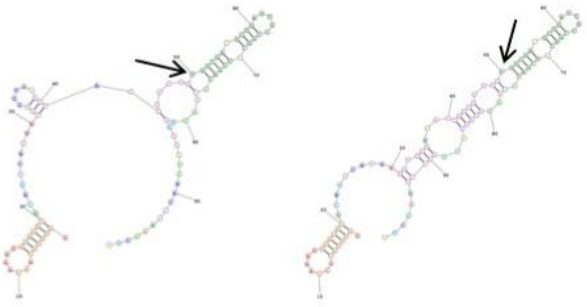
MaxExpect

Fold

YOL038W

Target UGUGU

GGUGUGCAACAAAGGUACACGCAGUCAGGAGGU
GUUAGACCAUUUGG**UGUGU**CGACGCUGAUUGCC
GGCUUCGACCCGAGAGAUGAUGAACCCAAGCUU

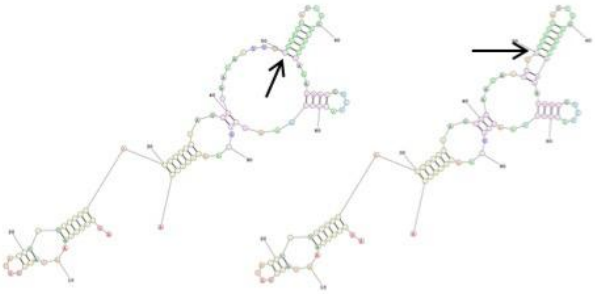


MaxExpect

Fold

Non-target UGUGU -1

AGGCGGUAAAGAGGGGUACCUGUGCUGUAGGUG
UCAAGGGUAAGAAU**UGUGU**AGUAUUGGGCUGCG
AAAGAAGGUCUACUUUGAAGCUGCAAGACACUA

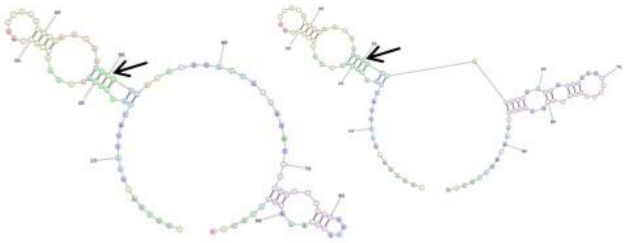


MaxExpect

Fold

Non-target UGUGU - 2

UGGAGAAGAAUUACGAUCGCAAAGAACCACCAGC
CACAGUGGAAGAA**UGUGU**CAAACUUACUGUAAG
AUCUCUGUUGGAGGUAGUUCAAACAGGUGCAA



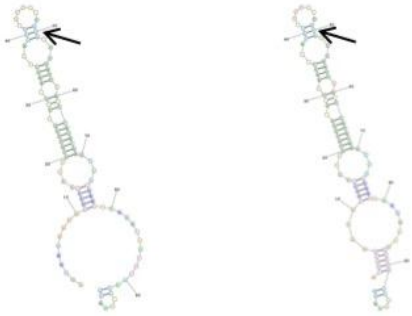
MaxExpect

Fold

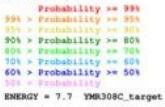
YMR308C

Target UGUAG

AUUGAAAAUAAUUCGCCAAUUGUGUGUGCUCAA
UCUAAUAUCUCCGC**UGUAG**UUGAUUCAGUCAUA
CAAGCCUUGAAUGAGAGAAGUUUGACCGAAAGG



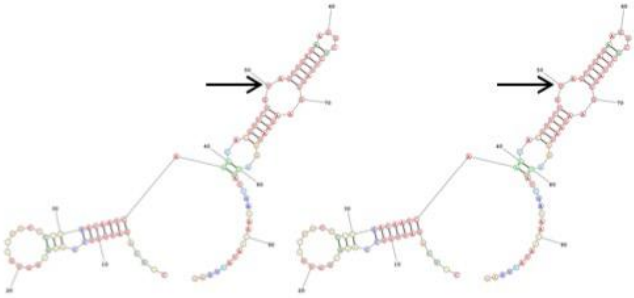
MaxExpect



Fold

Non-Target UGUAG

CUUCAGAUUGUCCAGGCGUUUGCUUCCCCUGAC
AAUCAAAUACGUUC**UGUAG**CUGAGAAGGCUCUU
AGUGAAGAAUGGAUUACCGAAAACAAUAUUGAG



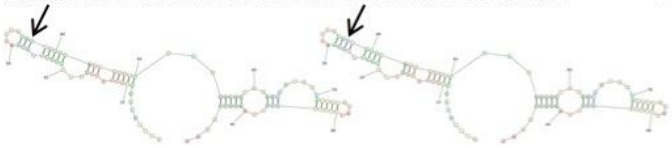
MaxExpect

Fold

YHR141C

Target UGUUA

AUGGGUAUGUGGACGAUUAGGAAUAGACAAACC
A**UGUUA**UUUAUCUCCAUUAGGGCGUGAGAGUGU
AAUUAGUACACAGGUACUACUAGAAUGCUAAAG



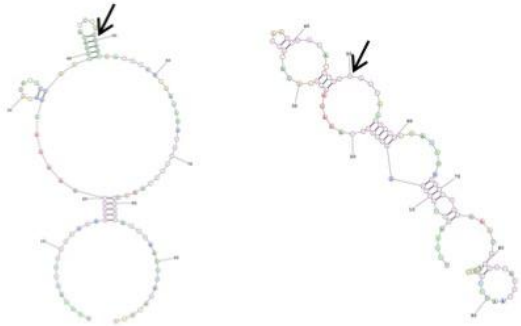
MaxExpect

Fold

YDR322C-A

Target UGUAG

UUGAAGCUGGUAGAGGAGGCAAAGAAGGAAUAC
GCCAAGCUACACCC**UGUAG**UAACUCCUAAAGAU
GUGCCUGCGAACGCCUCAUUUAAUUUGGAAGAU



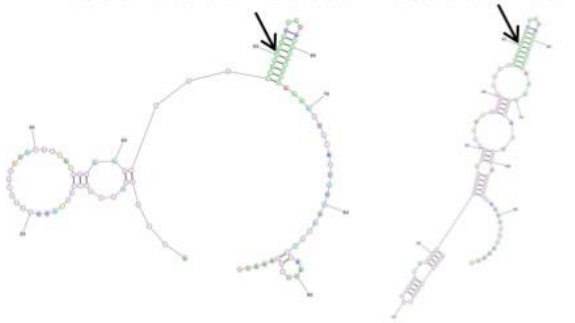
MaxExpect

Fold

YOR305W

Target AGUAG

ACCGGUGGCGCCCAUGAUGGUGGCGUAGACAUA
AAGGGAAGCUGGCC**AGUAG**AUGAUAUUUAUUGG
AAAAUUUCAUCGUUAAUGCCCAAUUUGGAAAUG

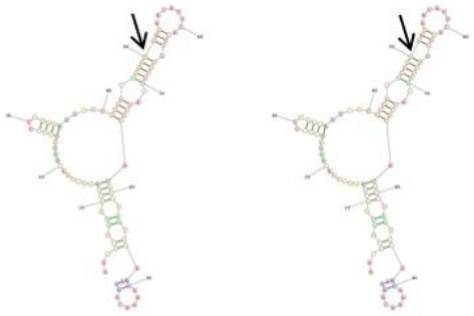


MaxExpect

Fold

Non-Target AGUAG

AAGGCUUUCACUAAAUCAAAAUUAUCUCCUAGAG
AAUUCCGUGAGUU**AGUAG**GAACAUUCACCUCAC
UAGUAUCACAUAGUCAACGCAACAAAACAGUA



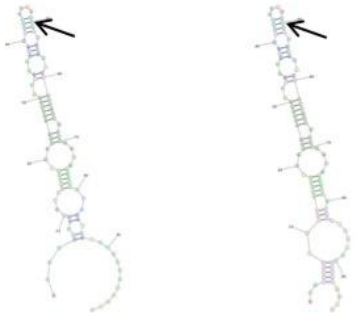
MaxExpect

Fold

YPR198W

Target UGUAG

AGAGUCGUUGCUGGGUUUGGAGGAAGUGGAAUU
GAAUCACUUGCUUU**UGUAG**UUGGAACAUCCAUU
GUCCGAGAAAACCAUAGAGGAAUUAUGAUAACG

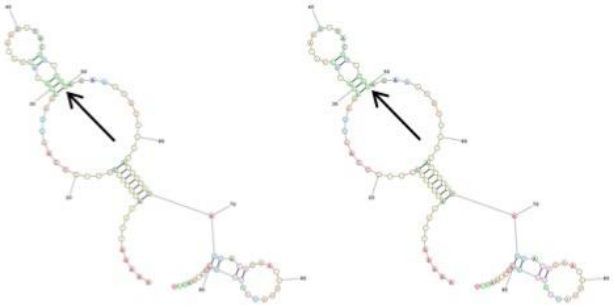


MaxExpect

Fold

Non-Target UGUAG

AAAAACCAGAUGCGCUCUUCCCAAUUAAACAUAC
CACAAUUUACGUC**UGUAG**AAACACUUUUAGCGU
AUAGCACAGAACAUUAUGAUGGCCCCCAAUCU



MaxExpect

Fold

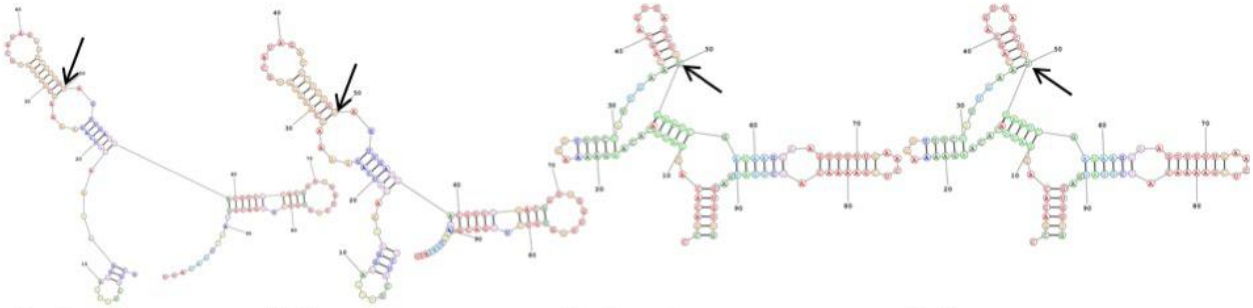
YGL115W

Target UGUAG

UCUACUAUUAUGGAUAACAUCAGAAAAGCAAGGG
UGCAUAGAUUCUU**UGUAG**UUGAUGACGUCGGAC
GGUUGGUUGGUGUCUUGACGUUAAGCGAUAUU

Non-Target UGUAG

CCGACACAGGAUUCACAAGAAAAGGUUUCUAUU
GAACAGCAGUUAGC**UGUAG**AAUCGAUAAGGAAG
UUUUUGAACUCGAAAACAUCUUAUGACGUGUUG



MaxExpect

Fold

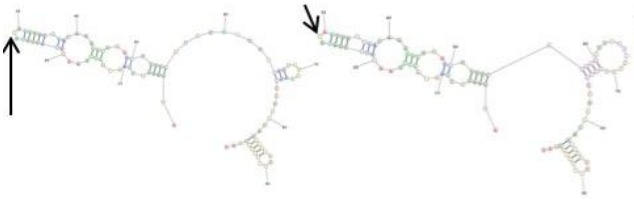
MaxExpect

Fold

YPR009W

Target UGUAA

AUGAAGCCGAACAAUCGAACUUGUGA**UGUAA**UU
ACCAACAAAGAUGAAUCUCUUCCUGCACUUUUG
CUGCCUGCACUGAACAGUUACACCUGUGACGAU



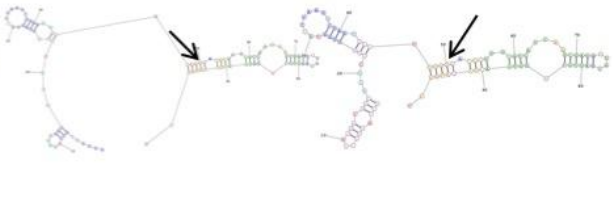
MaxExpect

Fold

YER150W

Target UGUAA

GCCAUGGCCUCUACGGCUCUCGGAUUGGUAUCU
AAUUCUAGUUCCUC**UGUAA**UCGUGGUACCAUCA
AGCGAUGCUACUAUUGCCGGUAACGAUACAGCC



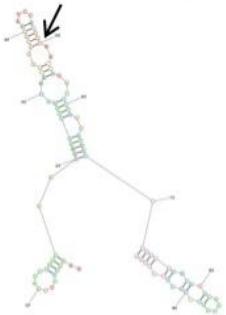
MaxExpect

Fold

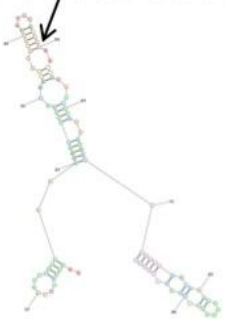
YLR354C

Target UGUAA

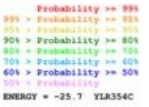
AAGGUAUUCAAGCUGCCAAAGAAUUGGAAGAAAA
GGACGGUAUCCAC**UGUAA**UUUGACUCUAUUAUU
CUCCUUCGUUCAAGCAGUUGCCUGUGCCGAGG



MaxExpect



Fold



YPL230W

Target UGUAG

ACUGCAGGCGUUCUUACGAGCAGCAAUGGUAAC
UUUGCCACCAAUAG**UGUAG**CGGCAUCAACUCCG
AAGAGGUCCAAAAGUGCUCGAAGGAAAACGUUC

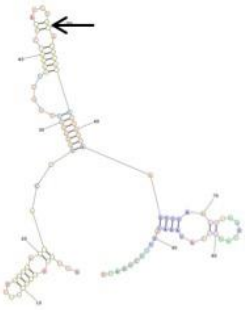
MaxExpect

Fold

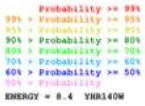
YHR140W

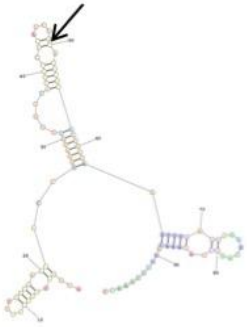
Target UGUAG

ACCGCGUGUCUUUUGACGAGCACCUGGGGGUUU
GUUAGGGCCACAUC**UGUAG**UCUUACCUCCAAGU
UUAAGUAAAGCAGGCCAUAAACAGUUUCUAACC

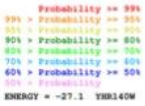


MaxExpect





Fold



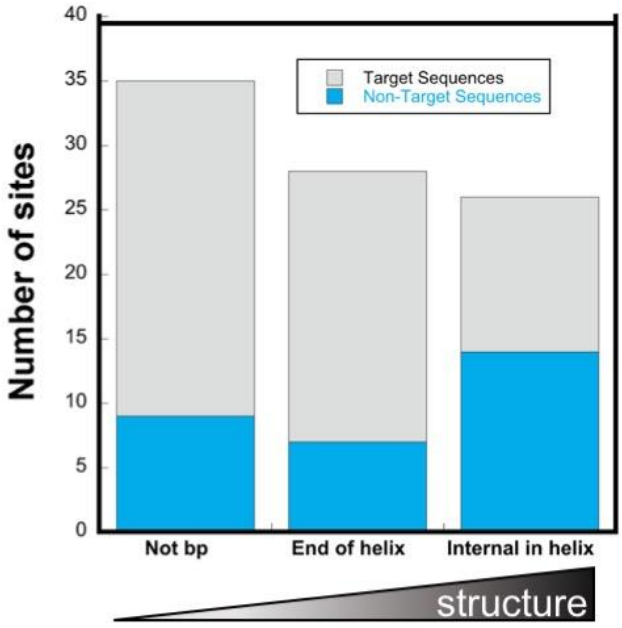


Figure S17: Model - Pus7 rapidly samples RNAs for specific modifiable sequences. RNAs contain multiple (often overlapping) potential Pus7 binding sites. These sites have varying degrees of accessibility to Pus due to their secondary/tertiary structures or occlusion by RNA-binding proteins. Pus7 rapidly samples all accessible sites on a given RNA, forming nonspecific interactions with most sequences. When Pus7 interacts with a modifiable (e.g. UGUAR) sequence, it forms a tighter, 'specific' interaction that results in Ψ installation. Only a handful of the potential Pus7 sites are modifiable and 'specific.'

Figure S18: Secondary structure predictions at 30°C and 45°C of randomly selected Pus7 heat shock targets Schwartz, et al. Cell (2014) (24).

YBL030C

Target UGUAG – 30°C **45°C**

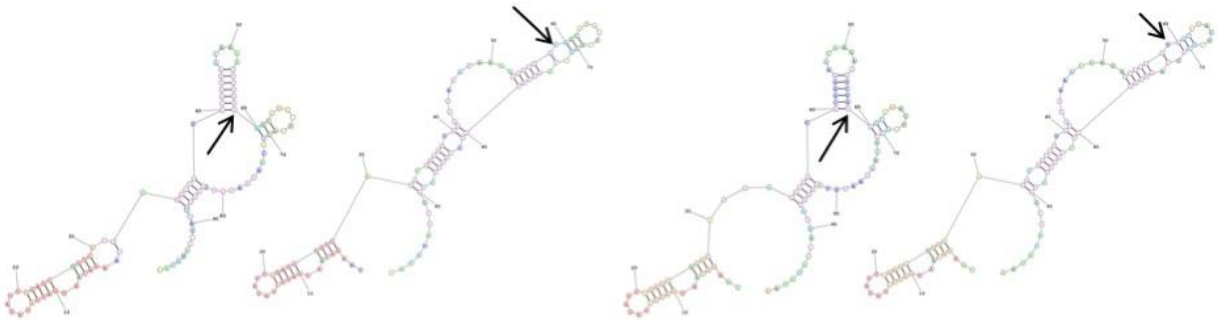
UACAAUGCUCUAAUGAAAAUAUUUUUCAUUGUAU
CUACCGCUUACAU**UGUAG**UGCUAUUACAAGGGU
CUAAAAGAACCAACACCAUUGCGUAUAAUGAA

MaxExpect Fold MaxExpect Fold

YBR103W

Target UGUAG – 30°C **45°C**

CAAGACAAUGAUUAUAUAUUUGAUCUGUCAUGGA
ACUGUGCGGGAAACAAGAUUUC**UGUAG**CAUACU
CGCUUCAAGAAGGUUCAGUUGUAGCCAUAUAA

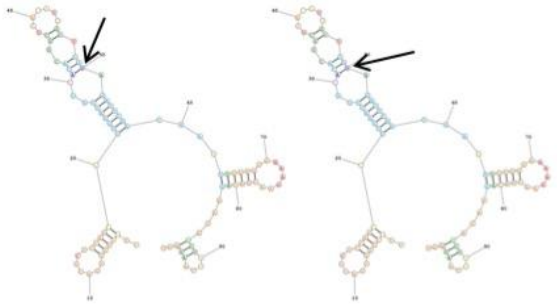


MaxExpect Fold MaxExpect Fold

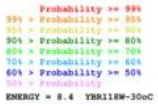
YBR118W

Target UGUAG – 30°C **45°C**

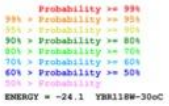
CUGCUGGUUACUCUCCAGUUUUGGAUUGUCACA
CUGCUCACAUUGCU**UGUAG**AUUCGACGAAUUGU
UGGAAAAGAACGACAGAAGAUCUGGUAAGAAGU

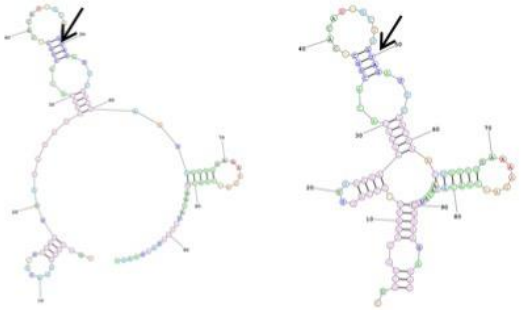


MaxExpect



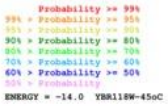
Fold





MaxExpect

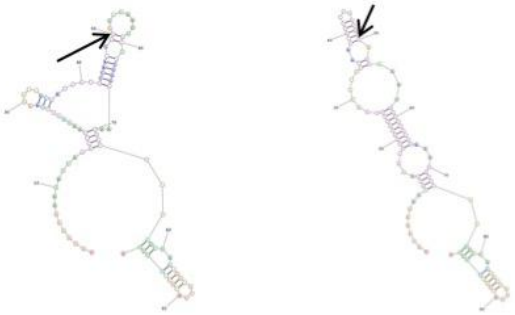
Fold



YBR191W

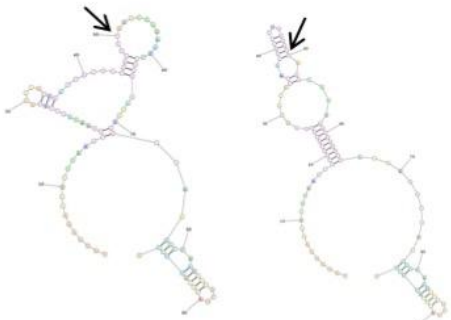
Target UGUAG – 30°C **45°C**

UAGAAAAAAGAUUAAACUUAAGAGUUGAACACAU
CAAGCACUCCAAG**UGUAG**ACAAGAAUUUUUGGAA
AGAGUUAAGGCCAAUGCUGCUAAGCGUGCUG



MaxExpect

Fold



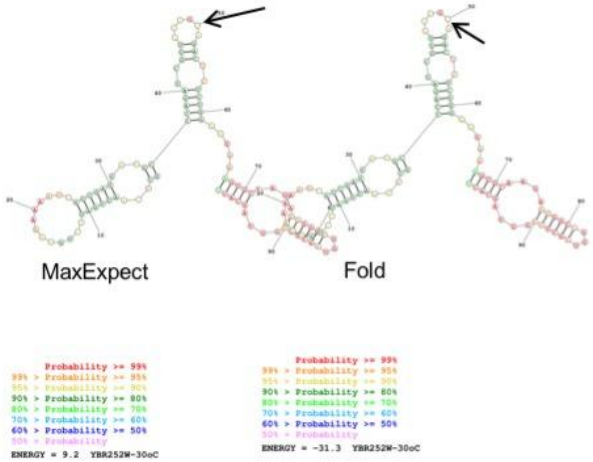
MaxExpect

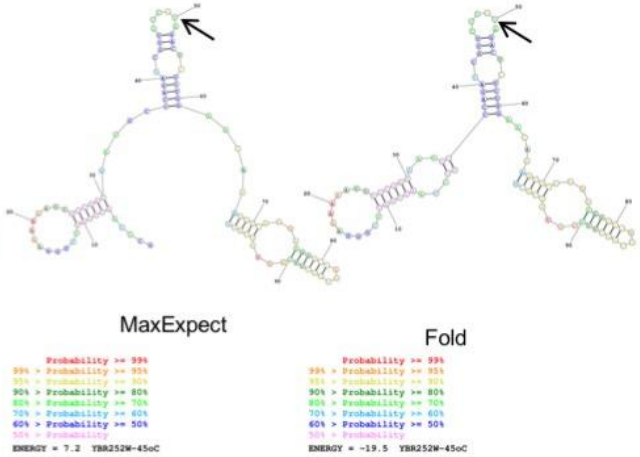
Fold

YBR252W

Target UGUAG – 30°C **45°C**

GCCCAAUUGAUUCUGGAAAAAAUUGUCGAUGAU
GCCCAGAUCGUUGU**UGUAG**ACUCUCUGGAAGAA
AGUGCAAGAGGGGCCGGUGGCUUUGGUAGCACU

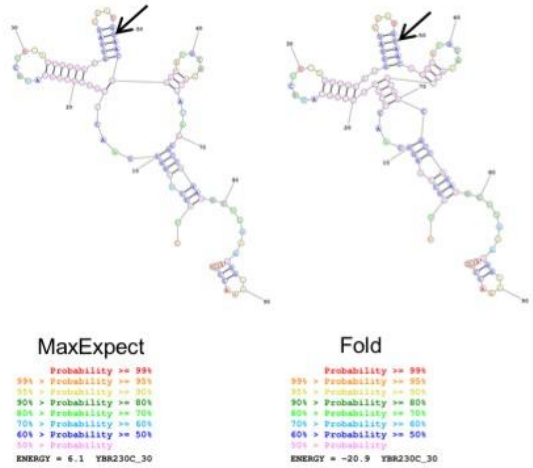


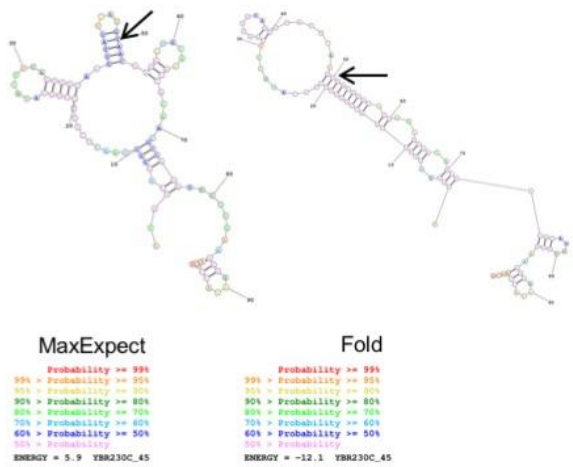


YBR230C

Target UGUAA – 30°C **45°C**

UUGUCUUGGUCAACUUGUUGCUGCAUUCUUCAG
UGGUAUGUUAUUUA**UGUAA**CGGGUAUGCGAACC
ACAACGCCAGAUUCUUGAAGGGGAAACCUAACU

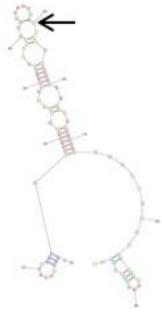




YDL022W – U403

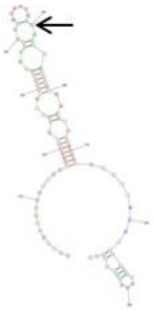
Target UGUAG – 30°C **45°C**

UCGACAUCAUCGUUUUCAACAUUCCACAUCAAUU
UUUGCCCCGUAUC**UGUAG**CCAAUUGAAAGGUCA
UGUUGAUUCACACGUCAGAGCUAUCUCCUGUC

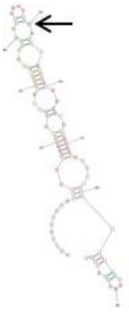


MaxExpect

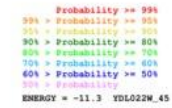
Fold



MaxExpect



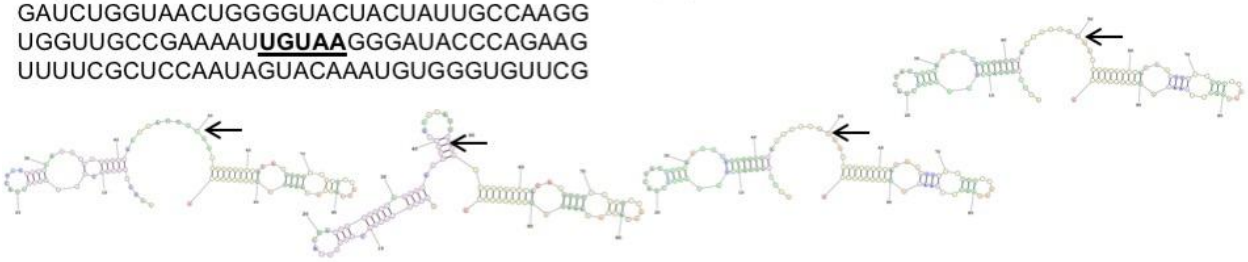
Fold



YDL022W – U172

Target UGUAA – 30°C **45°C**

GAUCUGGUAACUGGGGUACUACUAUUGCCAAGG
UGGUUGCCGAAAAU**UGUAA**GGGAUACCCAGAAG
UUUUCGCUCCAAUAGUACAAAUGUGGGUGUUCG



MaxExpect

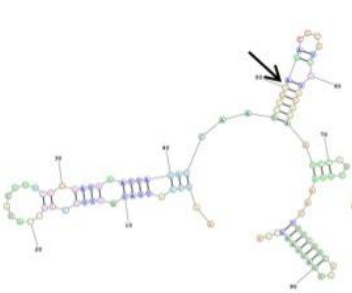
MaxExpect

Fold

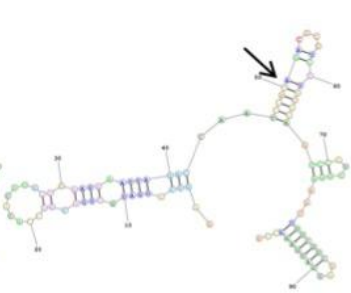
YDL084W

Target UGUAA – 30°C **45°C**

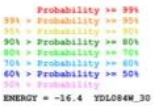
AAGUCAUGAUGUUUUCAGCCACACUUUCUCAAGA
AAUUAGACCAAUU**UGUAG**ACGCUUCUUACAGAAU
CCAUUGGAAAUUUUCGUCGAUGAUGAAGCUA

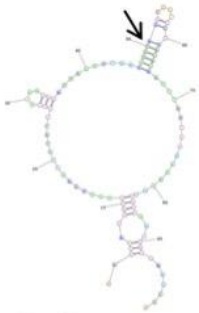


MaxExpect

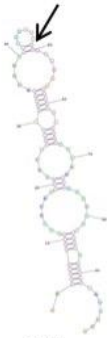


Fold





MaxExpect

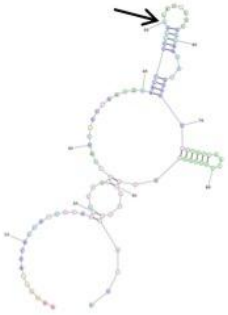


Fold

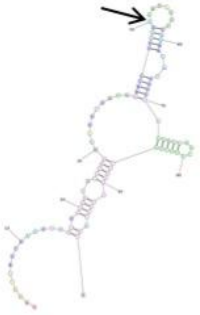
YDR037W

Target UGUAA – 30°C **45°C**

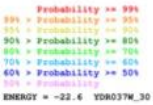
AAAAGGAUCAAGGUGAUGACGAAGCUCAAUUAGU
CGAUGAAACCUUC**UGUAA**UGCUCUAGAAUACGG
UUUACCACCAACUGGUGGUUGGGGUUGUGGUA

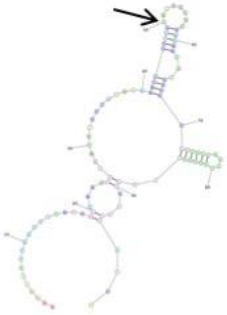


MaxExpect

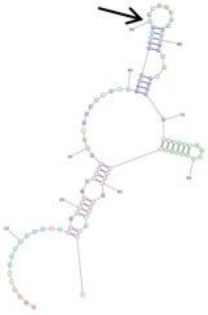


Fold

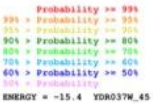




MaxExpect



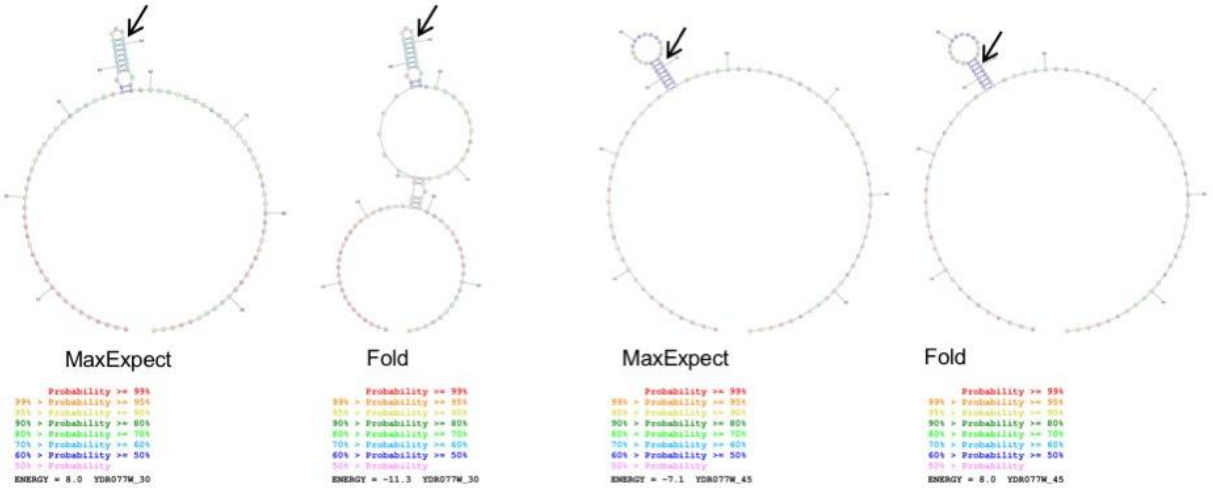
Fold



YDR077W

Target UGUAG – 30°C **45°C**

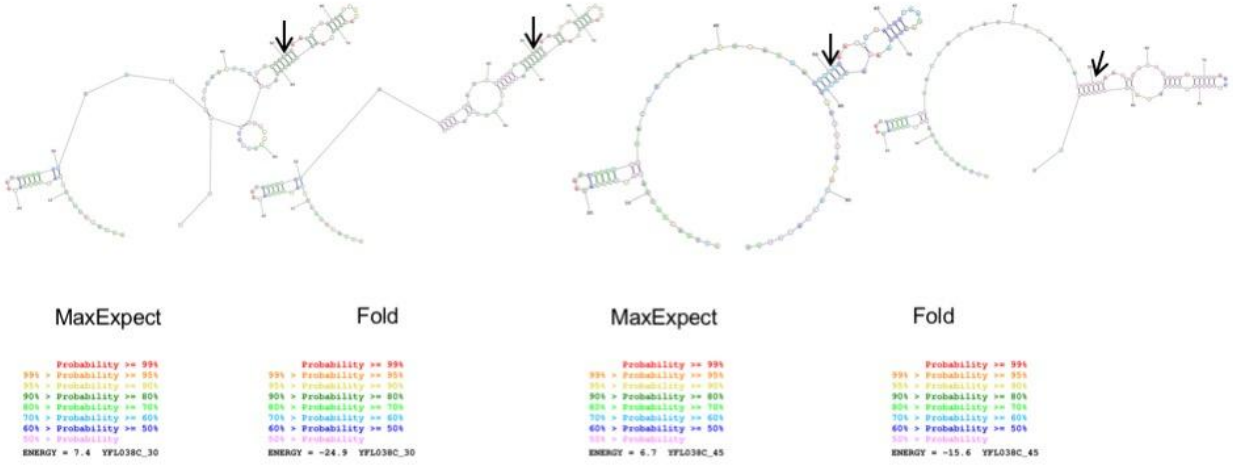
ACCACUCCUCCUUACAACCCAUCUACUGACUACA
CCACUGACUACAC**UGUAG**UCACUGAAUAUACUAC
UUACUGUCCAGAACCAACCACUUUCACCACA



YFL038C

Target UGUAG – 30°C **45°C**

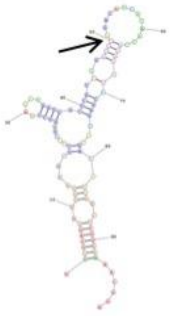
ACCAACGACUACAUCUCCACAAUUGGAGUGGACU
UCAAGAUUAAGAC**UGUAG**AACUGGACGGCAAGA
CUGUAAAGCUACAGAUUUGGGACACUGCAGGU

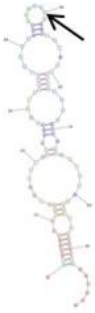


YGL116W

Target UGUAG – 30°C **45°C**

AUUGCAGAUCUGGACGUUCUUUUGGGCCUAAAU
GGUCGCUCGUCGGA**UGUAG**ACAUGACAACCACA
UUGCCGAGUUUGAAGCCACCUCUGCAAAACGGA



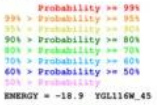


MaxExpect

Fold

MaxExpect

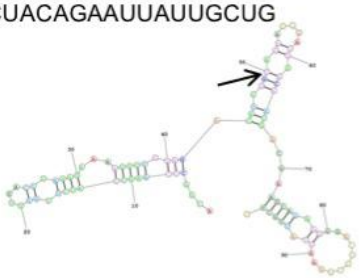
Fold

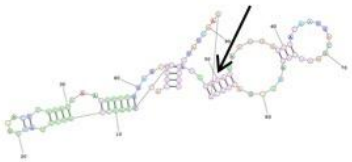


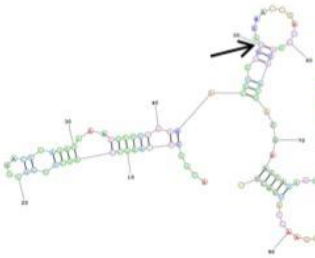
YIL076W

Target UGUAG – 30°C **45°C**

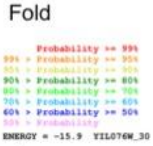
ACGGCUCAAGCUAUCUUGGGUGACUUAGAUAAA
AGUUUGGAGACAUG**UGUAG**AAGGGAUUGACAAU
GACGAAGCAGAAGGGACUACAGAAUUAUUGCUG

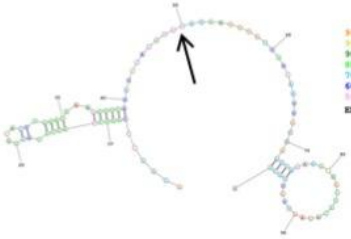






Fold





Fold

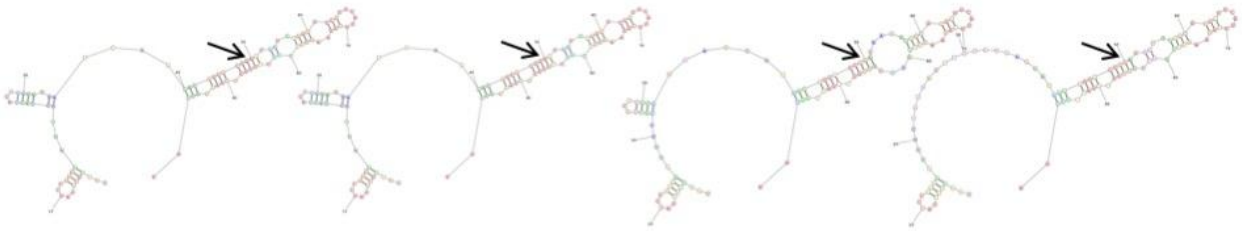
MaxExpect

MaxExpect

YJL158C

Target UGUAG – 30°C **45°C**

CAAUGUCUAUCCGGCAAUUUCUACAACUUGUAU
GAUCAAAACGUCGCCGAACAA**UGUAG**UGCCAUU
CAUUUGGAAGCUGUUUCUUUGGUCGACUGUUAA

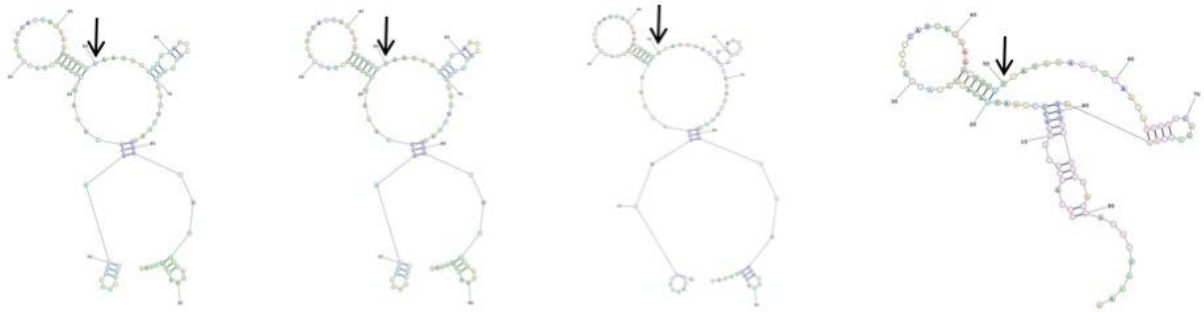


MaxExpect Fold MaxExpect Fold

YLR168C

Target UGUAG – 30°C **45°C**

UCCAGCAGGAAGCCCAAAUUACAGCAUAUGGAUC
CAUUAGAAAGCUG**UGUAA**UAAGAUGGAAGAUUG
GUCAGUUCAAAGGUUUUGCGAGAACGCUAAAA

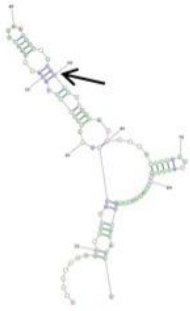


MaxExpect Fold MaxExpect Fold

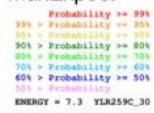
YLR259C

Target UGUAG – 30°C **45°C**

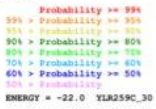
CCUUUAUUGCGUCGUGCUUACUCCUCUCAUAAA
GAAUUGAAAUUCGG**UGUAG**AAGGAAGAGCCUCC
CUUCUUAAGGGUGUCGAAACUUUAGCUGAAGCG

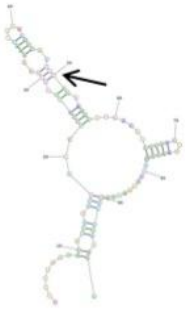


MaxExpect

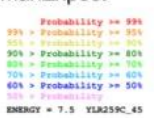


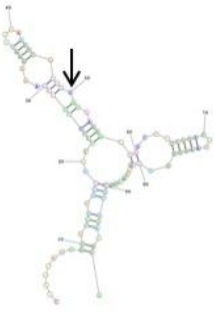
Fold





MaxExpect



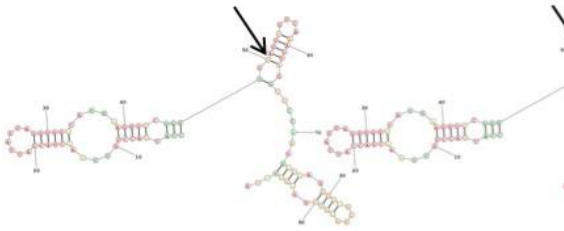


Fold

YMR083C

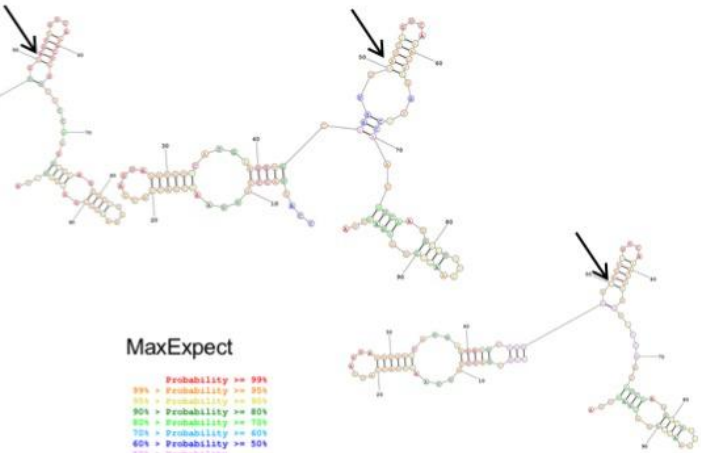
Target UGUAG – 30°C **45°C**

CCAUUACCUGUUAAACUACCAUUAGUAGGUGGU
CAUGAAGGUGCUGG**UGUAG**UUGUCAAACUAGGU
UCCAAUGUCAAGGGCUGGAAAGUCGGUGAUUUA



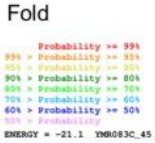
MaxExpect

Fold



MaxExpect

Fold



YMR226C

Target UGUAG – 30°C **45°C**

GACGUGGCUGAUCUGAUCGUCUAUGCAACUUCC
AGAAAACAAAAUAC**UGUAA**UUGCAGACACUUUAA
UCUUUCCAACAAACCAAGCGUCACCUCAUCAU

YNL098C

Target UGUAG – 30°C **45°C**

GGUGUUGGUAAAUCUGCUUUGACCAUACAAUUG
ACCCAAUCGCACUU**UGUAG**AUGAAUACGAUCCCA
CAAUUGAGGAUUCAUACAGGAAGCAAGUGGUG

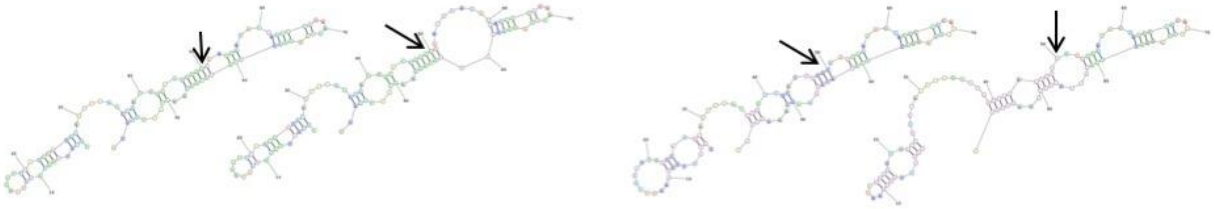


Figure S19: The observed rate constant for pseudouridinylation on short target 1 (ST1) is increased ~10-fold at elevated temperature. A. Observed rate constants for pseudouridinylation increase more than 10-fold as temperature increases, suggesting that increased conformational flexibility of the RNA structure allows more rapid access of PUS7. B. A set of stochastic structure predictions (32) demonstrating possible temperature-dependent changes in the structural environment of the target U in substrate sT1.

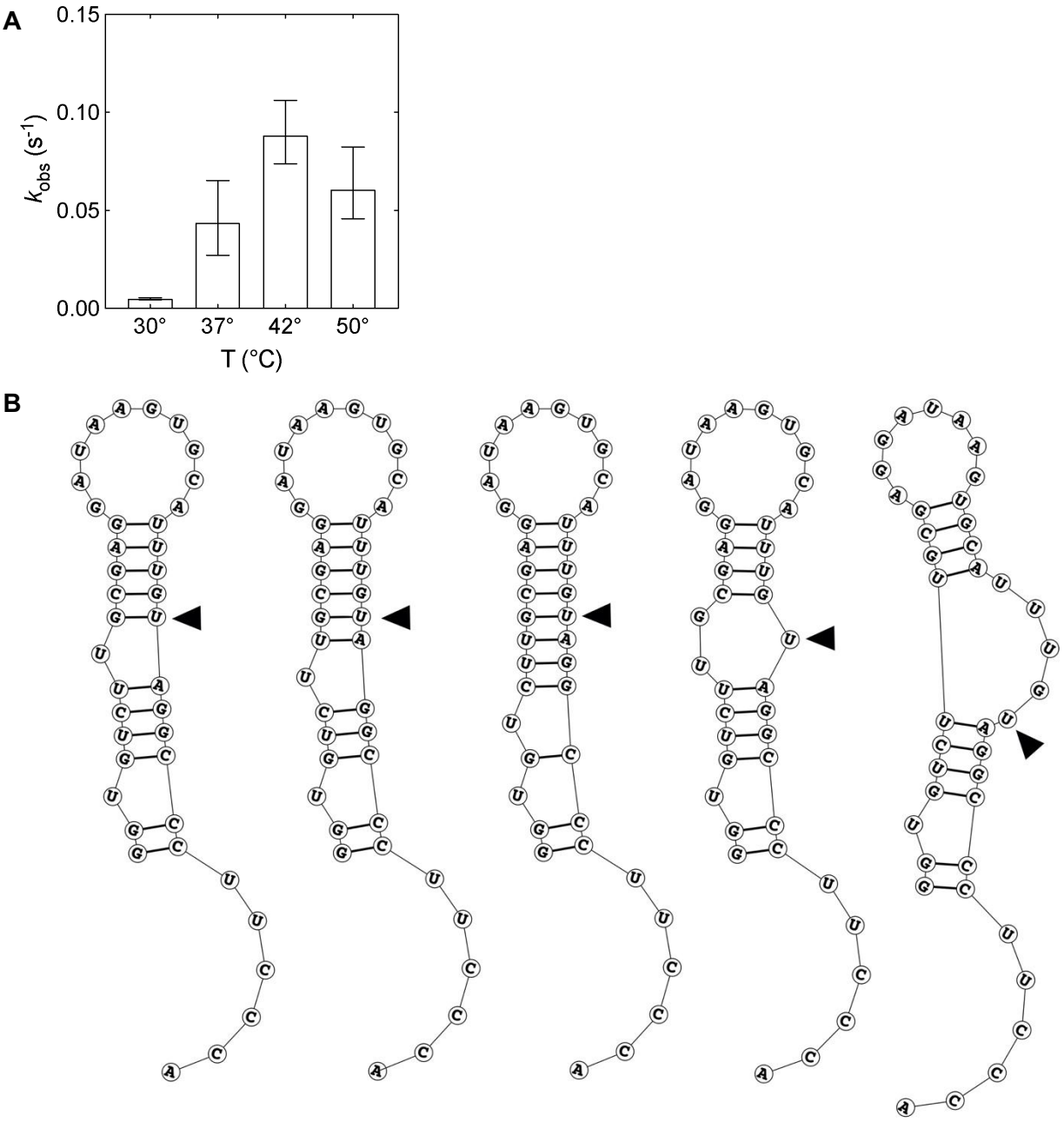
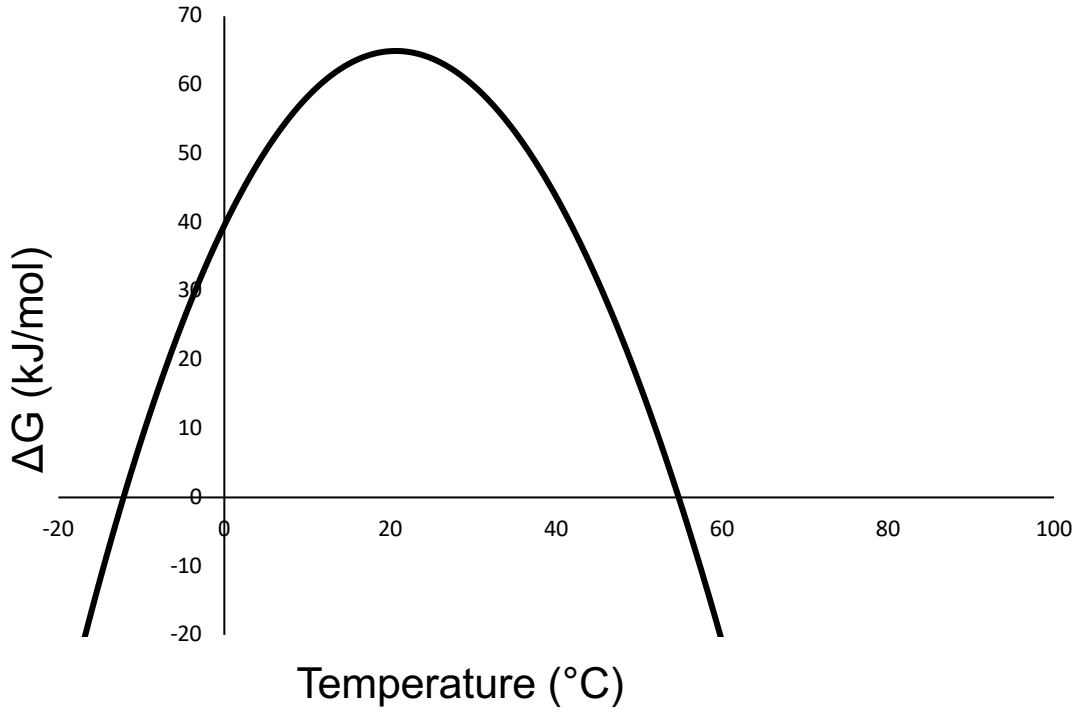


Figure S20: Modeled thermal stability of PUS7.
Using the chain length (N) of Pus 7, its stability was modeled as a function temperature range to find its maximum stability. Its maximum stability of about 65 kJ/mol is at approximately 22 ° C.



SUPPLEMENTAL APPENDIX TABLES

Table S1: Crystallographic parameters.

Table S1. X-Ray Crystallography Data Collection and Refinement Statistics

	Pus7
Data collection	
Beamline	APS, LSCAT 21-IDD
Wavelength (Å)	0.979
Temperature (K)	100
Resolution (Å)	48.27-3.20 (3.42-3.20)
Space group	*C*222
Cell dimensions (Å)	a = 117.9, b = 171.8, c = 105.3
Cell dimensions (°)	$\alpha = \beta = \gamma = 90$
Observed reflections	184,895 (31,045)
Unique reflections	18,019 (3,207)
R_{meas} (%)	17.8 (141.7)
R_{merge} (%)	17.8 (132.6)
$\langle I/\sigma \rangle$	9.5 (2.0)
CC(1/2)	0.996 (0.802)
Multiplicity	10.3 (9.7)
Completeness (%)	99.9 (100)
Overall *B* (Å^2) (Wilson plot)	121.9
Refinement	
Resolution range	46.32 - 3.20
Number of reflections (work/test set)	18017/881
R_{work}/R_{free} (%)	22.4/27.6
No. of non-H atoms	
Protein	9394
Water	14
Ligand	15
B-factors (Å^2)	
Protein	130.1
Water	88.7
Ligand	164.3
Rmsd deviations	
Bond lengths (Å)	0.0025
Bond angles (°)	1.21
Estimated coordinate error (Å); maximum likelihood based	0.4200
Cruickshank's DPI[1] (Å)	0.4688
Ramachandran plot	
Favored/allowed/outliers	87.7/12.1/0.2

MolProbity Score	1.63 (100^{th} percentile)
PDB	7MZV

Table S2: Impact of Pus7 mutations on CDC8 binding and modification

variant	k_{obs} (s^{-1}) [a,b]		k_{obs} defect (fold) [c]	$K_{D, app1}$ (nM) [d,e]	
WT[e]	9.9×10^{-1}	1.0×10^{-1}	1	76	15
D256A[e]		no reaction		60	16
K61A	2.6×10^{-2}	0.1×10^{-2}	38 6	400	200
F67A	4.6×10^{-3}	0.2×10^{-3}	210 30	180	40
E71A	5.2×10^{-3}	0.3×10^{-3}	190 30	210	50
F307Y	2.6×10^{-3}	0.1×10^{-3}	390 60	378	102
N305A	4.0×10^{-4}	$\leq 1 \times 10^{-5}$	2,400 300	230	60
F307A	1.3×10^{-5}	$\leq 1 \times 10^{-6}$	74,000 9,000	344	170
WT[f]	8.4×10^{-1}	0.5×10^{-1}	1	n.d.[g]	
H161A[f]	6.9×10^{-1}	0.9×10^{-1}	1.2 0.2	170	40
ΔID1[f]	3.8×10^{-1}	0.6×10^{-1}	2.2 0.5	160	40

[a] errors are standard error of the fit
[b] k_{obs} determined by fitting a curve of the form $y = 1 - e^{-k_{obs} \times t}$
[c] relative to WT at the same concentration on full-length cdc8 substrate
[d] K_D determined by curve fitting as described in Supplemental Appendix – Extended Methods
[e] All K_D values determined using D256A-double mutants, except for WT (no mutation), and the F307Y and D256A single mutants.
[f] Concentration of [Pus7] used to determine k_{obs} = 10 μM. All other reactions carried out with [Pus7] = 2 μM
[g] n.d., not determined

Table S3: Dissociation constants for Pus7 binding to various substrates.

substrate	variant	$K_{D,app1}$ (nM) [a,b]	
cdc8-A	WT	16	2
	D256A	57	4
cdc8-B	D256A	802	320
cdc8-C	WT	74	19
	D256A	131	13
sT1	D256A	34	4
	ΔID1-D256A	69	13
sT2	D256A	not analyzed – very weak	
	ΔID1-D256A	not analyzed – very, very weak	
$tRNA^{Asp}_{,GUC}$	D256A	16	1
	ΔID1-D256A	34	1

[a] errors are standard error of the fit
[b] K_D determined by curve fitting as described in the Extended Methods

Table S4: Observed rate constants for pseudouridinylation on different substrates.

substrate	variant	[PUS7] (μM)	k_{obs} (s^{-1}) [a,b]	
cdc8-A	WT	1	4.9×10^{-1}	0.5×10^{-1}
		5	9.3×10^{-1}	1.4×10^{-1}
		10	7.7×10^{-1}	0.8×10^{-1}
cdc8-B	WT	2	6.4×10^{-1}	1.0×10^{-1}
		10	7.8×10^{-1}	0.8×10^{-1}
cdc8-C	WT	2	8.9×10^{-1}	2.1×10^{-1}
		10	9.9×10^{-1}	1.4×10^{-1}
sT1	WT	10	5.1×10^{-3}	0.2×10^{-3}
	ΔID1	10	9.4×10^{-3}	0.7×10^{-3}
sT2	WT	10	2.4×10^{-4}	0.2×10^{-4}
$tRNA^{Asp,GUC}$	WT	10	9.1×10^{-3}	0.5×10^{-3}

[a] errors are standard error of the fit
[b] k_{obs} determined by fitting a curve of the form $fraction\ U \rightarrow = 1 - e^{-k_{obs} \times t}$

Table S5: RNAs used for biochemical assays.

substrate name	RNA sequence
cdc8-FL	GUCAAUCACGAUUGU**U**AGACGUUACUAAUAAGGGCAUUCAGGAAGU UGAAGCGCUUAUUUGG
CDC8-FL-NT	GUCAAUCACGAUUGCAGACGUUACUAAUAAGGGCAUUCAGGAAGU UGAAGCGCUUAUUUGG
CDC 8-A	GAUUGU**U**AGACGUUACUAAUAAGGGCAUUCAGGAAGUUGAAGCGCU UAUUUGG
CDC 8-A-NT	GAUUGCAGACGUUACUAAUAAGGGCAUUCAGGAAGUUGAAGCGCU UAUUUGG
CDC 8-B	GUCAAUCACGAUUGU**U**AGACGUUACU
CDC 8-B-NT	GUCAAUCACGAUUGCAGACGUUACU
CDC 8-C	GUCAAUCACGAUUGU**U**AGACGUUACUAAUAAGGGCGGAAGUGCGCU UAUUUGG
CDC 8-C-NT	GUCAAUCACGAUUGCAGACGUUACUAAUAAGGGCGGAAGUGCGCU UAUUUGG
ST1	GGUGUCUUGCGAGGAUAAGUGCAUUUGU**U**AGGCCCUUCCCA
SNT1	GGUGUCUUGCGAGGAUAAGUGCAUUUGCAGGCCCUUCCCA
ST2	GGGAUCUGU**U**AGCCCACCAA
SNT2	GGGAUCUGCAGCCCACCAA
tRNA^{Asp,GUC}	GCCGUGAUAG**U**UUAAUGGUCAGAAUGGGCGCUUGUCGCGUGCCA GAUCGGGGUUCAAUUCCCCGUCGCGGCGCCA
tRNA^{Asp,GUC}-NT	GCCGUGAUAGCUUAAUGGUCAGAAUGGGGCUUGUCGCGUGCCAG AUCGGGGUUCAAUUCCCCGUCGCGGCGCCA
CLAP-CDC8	GGCUAUUGGAUAAAGAGAUAAGGAAAGGCGAUGAGUCAAUCACGA UUGUAGACGUUACUAAUAAGGGCAUUCAGGAAGUUGAAGCGCUUA UUUGGCAAAUCGUUGAGCCUGUUUUGAGUACGCAUAU

SUPPLEMENTAL APPENDIX REFERENCES

1. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
2. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65(Pt 6):582-601.
3. Potterton L*, et al.* (2018) CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr D Struct Biol* 74(Pt 2):68-84.
4. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
5. Williams CJ*, et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* 27(1):293-315.
6. Anonymous (The PyMOL Molecular Graphics System (Schrödinger, LLC).
7. Milligan JF, Groebe DR, Witherell GW, & Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 15(21):8783-8798.
8. Wright JR, Keffer-Wilkes LC, Dobing SR, & Kothe U (2011) Pre-steady-state kinetic analysis of the three Escherichia coli pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis. *RNA* 17(12):2074-2084.
9. Kumar S, Stecher G, Li M, Knyaz C, & Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35(6):1547-1549.
10. Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7):1307-1320.
11. Felsenstein J (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 39(4):783-791.
12. Blanchet S*, et al.* (2018) Deciphering the reading of the genetic code by near-cognate tRNA. *Proc Natl Acad Sci U S A* 115(12):3018-3023.
13. Chou HJ, Donnard E, Gustafsson HT, Garber M, & Rando OJ (2017) Transcriptome-wide Analysis of Roles for tRNA Modifications in Translational Regulation. *Mol Cell* 68(5):978-992 e974.
14. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1).
15. Tardu M, Jones JD, Kennedy RT, Lin Q, & Koutmou KS (2019) Identification and quantification of modified nucleosides in Saccharomyces cerevisiae mRNAs. *ACS Chem Biol* 14(7):1403-1409.
16. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.
17. Kim D*, et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
18. Lauria F*, et al.* (2018) riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput Biol* 14(8):e1006169.
19. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
20. Jurrus E*, et al.* (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci* 27(1):112-128.
21. Hoang C & Ferre-D'Amare AR (2004) Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA* 10(7):1026-1033.

22. Dong A, Zeng, H., Walker, J.R., Bountra, C., Arrowsmith, C.H., Edwards, A.M., BROWN, P.J., WU, H., Structural Genomics Consortium (SGC) (2016) Crystal Structure of Human Pseudouridylate Synthase 7. (Protein Data Bank).
23. Carlile TM*, et al.* (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515(7525):143-146.
24. Schwartz S*, et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159(1):148-162.
25. Zhang W, *et al*. (2019) Sensitive and quantitative probing of pseudouridine modifications in mRNA and long noncoding RNA. *RNA* 25(9):1218-1225.
26. Myers, J. K.; Pace, C. N.; Scholtz, J. M. Denaturant m Values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. Protein Science : A Publication of the Protein Society 1995, 4 (10), 2138. https://doi.org/10.1002/pro.5560041020.
27. Ghosh, K.; Dill, K. A. Computing Protein Stabilities from Their Chain Lengths. PNAS 2009, 106 (26), 10649–10654. https://doi.org/10.1073/pnas.0903995106.
28. Sawle, L.; Ghosh, K. How Do Thermophilic Proteins and Proteomes Withstand High Temperature? Biophys J 2011, 101 (1), 217–227. https://doi.org/10.1016/j.bpj.2011.05.059.
29. Watson, M. D.; Monroe, J.; Raleigh, D. P. Size-Dependent Relationships between Protein Stability and Thermal Unfolding Temperature Have Important Implications for Analysis of Protein Energetics and High-Throughput Assays of Protein–Ligand Interactions. J. Phys. Chem. B 2018, 122 (21), 5278–5285. https://doi.org/10.1021/acs.jpcb.7b05684.
30. Rees, D. C.; Robertson, A. D. Some Thermodynamic Implications for the Thermostability of Proteins. Protein Sci 2001, 10 (6), 1187–1194.
31. Robertson, A. D.; Murphy, K. P. Protein Structure and the Energetics of Protein Stability. Chem. Rev. 1997, 97 (5), 1251–1268. https://doi.org/10.1021/cr960383c.
32. Harmanci, A. O., Sharma, G., & Mathews, D. H. (2009). Stochastic sampling of the RNA structural alignment space. Nucleic Acids Research. 37, 4063-4075.